

# 图像特征提取与检索技术

孙君顶 等著

电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

## 内 容 简 介

本书对基于内容的图像检索 (Content-Based Image Retrieval, CBIR) 技术的基本原理、图像特征提取与检索方法进行了比较详细的介绍和讨论, 并融入了作者多年来的相关研究成果。本书共有 6 章, 第 1 章介绍了 CBIR 的发展与现状、研究内容及涉及的关键技术, 第 2 章介绍了图像低层特征的提取与表达技术, 第 3 章介绍了基于压缩域的图像检索技术, 第 4 章介绍了视觉注意计算模型, 第 5 章介绍了自动图像标注技术, 第 6 章介绍了子空间特征提取技术。

本书层次分明, 内容翔实, 理论分析与算法实践相结合, 力求实用。本书既可作为高等院校计算机科学、信号和信息处理、图书情报等相关专业研究生的教材, 也可作为广大从事模式识别、多媒体分析、信息检索等研究、应用和开发工作的科技工作者和高等院校师生的科研参考书。

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究。

## 图书在版编目 (CIP) 数据

图像特征提取与检索技术 / 孙君顶等著. —北京: 电子工业出版社, 2015.7

ISBN 978-7-121-25271-6

I. ①图… II. ①孙… III. ①图象数据库—情报检索 IV. ①G354.49

中国版本图书馆 CIP 数据核字 (2014) 第 303395 号

策划编辑: 董亚峰

责任编辑: 韩玉宏

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1 092 1/16 印张: 23.5 字数: 602 千字

版 次: 2015 年 7 月第 1 版

印 次: 2015 年 7 月第 1 次印刷

定 价: 59.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 [zlts@phei.com.cn](mailto:zlts@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

服务热线: (010) 88258888。

# 前言

---

随着数字影像技术和网络技术的迅速发展，数字图像已经成为一种被广泛使用的媒体形式。相比于文本信息，数字图像更为形象生动、逼真直观，是一种独立性很强的信息载体，所以被迅速地应用于各种领域，如数字图书馆、新闻媒体、医学图像管理、卫星遥感图像、商标版权管理及地理信息系统等。

但是，目前数字化设备还无法实现对大量图像库的有效管理，用户还需要根据特定的分类技术和检索技术实现管理和查询，因此建立快速、有效的图像检索系统已经成为一个重要的研究方向。图像检索系统不仅需要能够驾驭巨大的图像库，快速响应用户的查询要求，而且还应当能够反馈给用户准确的、尽可能多的图像资料。为了实现上述目标，研究者进行了大量的研究，形成了具有指导意义的理论，提出了很多具有一定效果的图像检索技术和系统。从发展的过程来看，图像检索技术主要包括基于文本的图像检索技术、基于内容的图像检索技术及两者的融合技术——自动图像标注技术。

国内外至今已经出现了大量有关基于内容的图像检索技术方面的文章和著作，国际上每年也都召开许多有关信息检索技术方面的学术会议，许多大会都有图像检索技术的主题和分会。早期的著作《基于内容的视觉信息检索》（章毓晋著，科学出版社 2003 年出版）及《网上多媒体信息分析与检索》（庄越挺等著，清华大学出版社 2002 年出版）对基于内容的图像检索技术作了一定的阐述；2007 年，清华大学出版社出版了周明全等著的《基于内容图像检索技术》，对基于内容的图像检索技术作了较全面的论述；我们于 2009 年在电子工业出版社出版的著作《图像低层特征提取与检索技术》主要针对图像低层特征的提取与检索进行了全面的论述。近年来，图像检索技术发展迅速，各种新技术及方法不断涌现，我们在传统图像检索技术的基础上，进一步对图像检索新技术进行了总结。

全书共有 6 章，具体内容安排如下：第 1 章介绍了 CBIR 的发展与现状、CBIR 的研究内容与其所涉及的关键技术、CBIR 的应用及一些经典的 CBIR 系统；第 2 章介绍了图像低层特征的提取与表达技术，主要涉及图像的颜色、形状和纹理 3 种基本特征，又介绍了 MPEG-7 中的图像低层特征描述符；第 3 章介绍了基于压缩域的图像检索技术，包括空间压缩域和变换压缩域中常用的描述算法，并介绍了两种基于 DCT 压缩域的图像纹理及形状特征提取方法；第 4 章介绍了视觉注意计算模型，引入了基于特征加权、基于高斯混合和基于 CIE Lab 的 3 种视觉注意计算模型；第 5 章介绍了自动图像标注技术，主要讨论了图像视觉特征选择、低层特征到高层语义之间映射模型的建立两个方面的问题；第 6 章针对图像检索中的维数灾难问题，详细讨论了子空间特征提取技术。

本书由河南理工大学孙君顶博士撰写第 1 章、毋小省副教授撰写第 2 章、赵珊博士撰写第 3 章、郭海儒博士撰写第 4 章、王科平博士撰写第 5 章、王永茂博士撰写第 6 章，全书由孙君顶博士负责统稿及审校。本书结合 CBIR 技术的研究现状及发展方向，既参考了许多他人的有关文献，也结合了作者近年来在该领域的研究成果。

本书的出版得到河南省骨干教师资助计划（2010GGJS-059）、河南省国际合作项目（134300510057）及河南理工大学计算机学院的资助。

由于作者水平有限及国内外针对 CBIR 技术研究的逐步深入，书中的不妥与疏漏之处在所难免，敬请读者指正。

作者

2015 年 6 月



# 目 录

第 1 章 基于内容的图像检索与关键技术 .....	1
1.1 图像检索技术的发展 .....	1
1.1.1 基于文本的图像检索 .....	2
1.1.2 基于内容的图像检索 .....	3
1.1.3 自动图像标注技术 .....	6
1.1.4 国内外研究状况 .....	6
1.2 CBIR 的研究内容 .....	10
1.2.1 特征提取与匹配 .....	10
1.2.2 索引机制 .....	10
1.2.3 用户接口 .....	11
1.3 CBIR 的关键技术 .....	12
1.3.1 基本检索原理 .....	12
1.3.2 图像内容及检索层次 .....	13
1.3.3 常用特征描述方法 .....	14
1.3.4 特征匹配技术 .....	19
1.3.5 稀疏表示技术 .....	25
1.3.6 性能评价准则 .....	27
1.4 CBIR 的应用与经典系统 .....	30
1.4.1 CBIR 的应用 .....	30
1.4.2 经典 CBIR 系统介绍 .....	31
1.5 本书内容安排 .....	38
参考文献 .....	39

第 2 章 图像低层特征的提取与表达 .....	45
2.1 颜色特征的提取与表达 .....	45
2.1.1 颜色空间 .....	45
2.1.2 颜色量化 .....	50
2.1.3 全局颜色特征 .....	51
2.1.4 空间颜色特征 .....	56
2.2 形状特征的提取与表达 .....	68
2.2.1 概述 .....	68
2.2.2 基于轮廓的描述方法 .....	69
2.2.3 基于区域的描述方法 .....	89
2.3 纹理特征的提取与表达 .....	103
2.3.1 概述 .....	103
2.3.2 常用的纹理分析方法 .....	104
2.3.3 局部二值模式 .....	116
2.3.4 纹理基元共生矩阵 .....	128
2.4 MPEG-7 中的图像特征描述符 .....	131
2.4.1 颜色描述符 .....	133
2.4.2 形状描述符 .....	134
2.4.3 纹理描述符 .....	135
参考文献 .....	136
第 3 章 基于压缩域的图像检索技术 .....	146
3.1 概述 .....	146
3.1.1 图像压缩技术 .....	147
3.1.2 静态图像压缩标准 .....	153
3.1.3 压缩域图像检索的原理 .....	162
3.1.4 压缩域图像检索的研究内容 .....	164
3.1.5 压缩域图像检索的研究方法 .....	164
3.2 空间压缩域技术 .....	166
3.2.1 矢量量化 .....	166
3.2.2 分形编码 .....	169
3.2.3 预测编码 .....	171
3.3 变换压缩域技术 .....	172
3.3.1 基于 DFT 压缩域 .....	172
3.3.2 基于 DCT 压缩域 .....	173

3.3.3 基于小波压缩域 .....	181
3.3.4 基于 K-L 变换域 .....	186
3.4 空间域和变换域的融合检索 .....	188
3.5 DCT 压缩域内的纹理特征 .....	189
3.5.1 复杂度的定义 .....	190
3.5.2 复杂度直方图 .....	191
3.6 DCT 压缩域内的形状特征 .....	193
3.6.1 理想边缘模型 DCT 块的分类 .....	193
3.6.2 空间边缘分布特征的提取 .....	195
参考文献 .....	196
第 4 章 视觉注意计算模型 .....	205
4.1 概述 .....	205
4.1.1 人类视觉系统 .....	205
4.1.2 视觉系统理论 .....	207
4.1.3 研究现状 .....	214
4.2 基于特征加权的视觉注意计算模型 .....	219
4.2.1 模型实现过程 .....	219
4.2.2 物体识别实验 .....	223
4.2.3 物体搜索实验 .....	226
4.3 基于高斯混合的视觉注意计算模型 .....	229
4.3.1 高斯混合模型 .....	230
4.3.2 基于 GMM 的视觉注意计算模型 .....	232
4.3.3 实验与分析 .....	236
4.4 基于 CIELab 的视觉注意计算模型 .....	239
4.4.1 模型实现过程 .....	240
4.4.2 实验与分析 .....	245
参考文献 .....	255
第 5 章 自动图像标注技术 .....	261
5.1 概述 .....	261
5.1.1 自动图像标注概述及研究意义 .....	261
5.1.2 自动图像标注的关键问题 .....	264
5.2 图像视觉特征选择 .....	265
5.2.1 视觉特征选择 .....	265
5.2.2 视觉特征加权 .....	266

5.3 自动图像标注模型	273
5.3.1 基于生成模型的标注方法	273
5.3.2 基于判别模型的标注方法	279
5.3.3 基于多示例学习的标注方法	289
参考文献	314
<b>第 6 章 子空间特征提取技术</b>	<b>321</b>
6.1 概述	321
6.1.1 降维原因	321
6.1.2 子空间特征提取方法的形式化描述及分类	323
6.2 经典的子空间特征提取方法	324
6.2.1 线性方法	324
6.2.2 核方法	326
6.2.3 流形方法	328
6.2.4 半监督方法	333
6.2.5 张量方法	334
6.2.6 图嵌入框架	334
6.3 基于自适应近邻图嵌入的局部鉴别投影方法	339
6.3.1 方法提出的背景	339
6.3.2 LFDA	339
6.3.3 LADP	342
6.4 基于对角图像的模糊线性鉴别分析	347
6.4.1 方法提出的背景	347
6.4.2 FLDA	347
6.4.3 对角图像	353
6.4.4 DiaFLDA	354
6.5 DCT 域内拉普拉斯值排序的子空间特征提取方法	357
6.5.1 方法提出的背景	357
6.5.2 离散余弦变换 (DCT)	357
6.5.3 局部保持能力判据	359
6.5.4 DCT/LS+LPP	361
参考文献	362

## 基于内容的图像检索与关键技术

近年来，基于内容的图像检索（Content-Based Image Retrieval, CBIR）技术得到了广泛关注和深入研究，本章主要介绍 CBIR 技术的发展历程、研究内容、关键技术及 CBIR 的应用与经典系统。

### 1.1 图像检索技术的发展

随着数字影像技术和网络技术的迅速发展，数字图像已经成为一种被广泛使用的媒体形式。相比于文本信息，数字图像更为形象生动，更逼真直观，是一种独立性很强的信息载体，所以被迅速地应用于各种领域，如数字图书馆、新闻媒体、医学图像管理、卫星遥感图像、商标版权管理及地理信息系统等。

但是，目前数字化设备还无法实现对大量图像库的有效管理，用户还需要根据特定的分类技术和检索技术实现管理和查询，因此建立快速、有效的图像检索系统已经成为一个重要的研究方向。图像检索系统不仅需要能够驾驭巨大的图像库，快速响应用户的查询要求，而且还应当能够反馈给用户准确的、尽可能多的图像资料。为了实现上述目标，研究者做了大量的研究，形成了具有指导意义的理论，提出了很多具有一定效果的图像检索技术和系统。

从图像检索技术发展的过程来看，主要包括基于文本的图像检索、基于内容的图像检索及二者的融合技术——自动图像标注技术<sup>[1,2]</sup>。

### 1.1.1 基于文本的图像检索

基于文本的图像检索（Text-Based Image Retrieval, TBIR）技术的历史可以追溯到 20 世纪 70 年代末期。当时流行的图像检索技术是将图像作为数据库中存储的一个对象，用关键字或自由文本对其进行描述，查询操作是基于该图像的文本描述进行精确匹配或概率匹配，因此这种图像检索技术实质上是采用文本检索技术实现对图像的检索。

TBIR 技术在图像检索中得到了广泛应用，Google、Baidu 等早期的搜索引擎均采用这种方式来检索图像。图 1.1（a）及图 1.1（b）分别给出了 Google 和 Baidu 早期两个搜索引擎的一次检索结果示例，搜索的文本为“鲜花”。从这两大搜索引擎的检索结果中我们也可了解该搜索技术的特点，两个搜索引擎都是将标注文本中含有“鲜花”两字的图片检索了出来。



(a) Google 针对“鲜花”的检索结果

图 1.1 TBIR 示例



(b) Baidu 针对“鲜花”的检索结果

图 1.1 TBIR 示例 (续)

然而，TBIR 技术需要人工提前对图像库中的图像进行归纳和注释，图像检索结果也完全依赖于人工标注信息。该技术存在以下几个无法解决的问题。

(1) 每一幅图像都需要人工进行注释，对目前海量的图像数据来说，完全采用人工的方法都会遇到难以克服的困难。

(2) 人工注释具有很强的主观性，对于同一幅图像，不同人可能有着不同的看法，因此标注信息会不尽一致；而且，一旦标注后就很难更新和改变，这使得在很多情况下文本标注并不能满足实际需求。

(3) 一幅图像所包含的意义非常丰富，“百闻不如一见”、“一图值千言”都说明了这个事实，而人工注释的少量文字很难充分表达一幅图像的内涵。

(4) 不同国家、不同民族很难用同一种语言对图像加注标识，而且对图像语义理解的差异也很大，因此不可能形成一种统一检索方法。

### 1.1.2 基于内容的图像检索

区别于原有系统对图像进行人工标注的做法，CBIR 技术自动提取每幅图像的视觉内容特征作为索引（包括直接从被压缩的图像中提取特征），如颜色、纹理、形状等，然后通过计算比较这些特征和查询条件之间的距离，来决定两幅图像的相似程

度。这里，图像内容的描述及提取不再依赖于人的手工标注，而是借助于从图像中自动提取的视觉特征，检索过程也不再是关键字匹配，而是视觉特征间的相似匹配。该技术的研究涉及人工智能、计算机视觉、信号处理、模式识别、认知心理学、数据库、人机交互等诸多学科领域，具有重要的理论意义。CBIR 技术具有以下几个特点。

(1) 基于内容的图像检索突破了传统的基于表达式检索的局限，它直接对图像进行分析和抽取特征，并利用这些描述图像内容的特征来建立索引。

(2) 基于内容的图像检索实质上是一种近似匹配的技术。在检索的过程中，它采用某种相似性度量对图像库中的图像进行匹配，以获得查询结果。这一点与常规数据库检索的精确匹配方法有明显不同。

(3) 特征提取和索引建立可由计算机自动实现，避免了人工描述的主观性，也大大减少了工作量。

(4) 针对 CBIR 中存在的低层特征和上层理解之间的语义鸿沟问题，研究者可以采用相关反馈、机器学习、图像分割、基于视觉注意度等手段，提高检索结果与用户满意度的匹配程度。

CBIR 的一般框架如图 1.2 所示。

目前，Google 和 Baidu 两个图片搜索引擎都实现了基于内容的图像检索。图 1.3 给出了一个示例，图 (a) 所示为输入的查询图像，图 (b) 和图 (c) 给出了这两个搜索引擎的基于示例查询的检索结果。

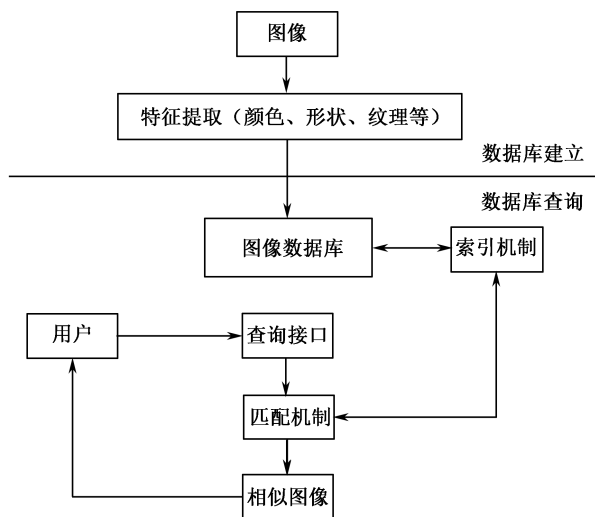


图 1.2 CBIR 的一般框架

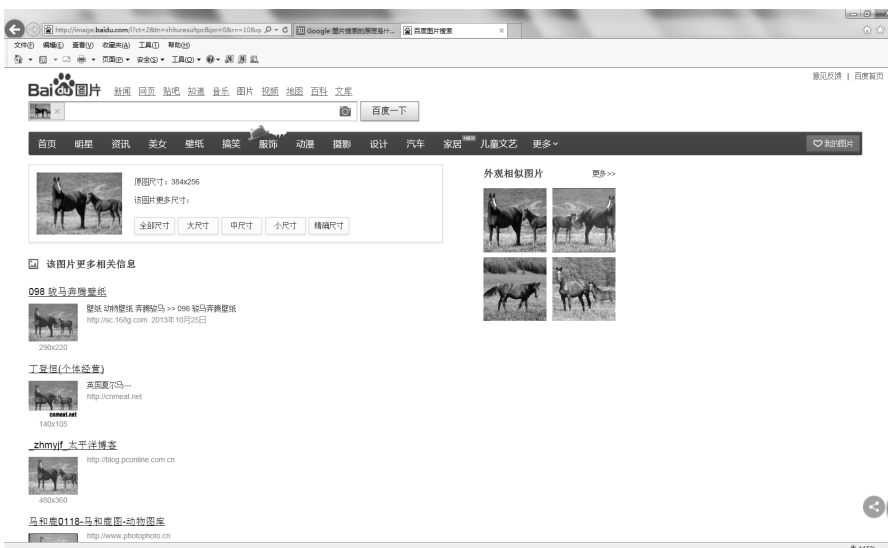




(a) 示例图像



(b) Google 检索结果



(c) Baidu 检索结果

图 1.3 CBIR 示例

### 1.1.3 自动图像标注技术

---

由于人类所理解的图像与用低层视觉特征来表达的图像之间存在着很大的差距，即在图像语义和视觉特征之间横亘着语义鸿沟。为了实现更为贴近用户理解能力的、用自然语言描述的查询方式，对图像语义标注的研究逐渐引起人们的广泛注意。

自动图像标注 (automatic image annotation) 就是根据图像低层视觉特征，使用语义关键字或标签来表示一幅图像的语义内容，进而可以将图像检索转化为基于文本的检索。该技术将基于文本的图像检索技术与基于内容的图像检索技术有效地结合起来。随着图像处理、机器学习和自然语言处理等技术的发展，近几年出现了许多图像语义的自动标注模型，其核心在于从已有训练数据中对高级语义概念与图像的低级视觉特征的关系进行自动建模，从而使用学习到的模型对新的图像进行标注。因此，图像语义的自动标注可以有效避免基于人工标注的图像检索系统所面临的一系列问题，使得大规模图像基于语义检索应用更具现实性。

### 1.1.4 国内外研究状况

---

基于内容的图像检索技术始于 20 世纪 90 年代初期，由于此项技术涉及的领域很多，以及该技术对信息检索领域的重要作用，因此迅速成为研究热点。各大研究机构、公司和高校，如 IBM、MIT、Columbia、Stanford、UIUC 等都对该技术进行了深入研究，并推出了各自的图像检索系统。

在国内，各大重点院校和科研机构都开展了关于基于内容的图像、视频、音频的检索方法的研究。清华大学结合国家 863 高技术研究发展项目，开展了“Web 上基于内容的图像检索”的研究，研究目标是开发能在 Internet/Intranet 环境下，通过友好的人机界面，利用主颜色、纹理、颜色分布和轮廓等图像特征或样本进行图像检索的方法和工具。中国科学院计算技术研究所和北京图书馆联合开发了“基于特征的多媒体信息检索系统 MIRES”。浙江大学开发了 Photo Navigator、Photo Engine 和 WebscopeCBR 等系统。总之，目前越来越多的科研机构都加入了研究基于内容的图像检索技术的行列中来。

目前，CBIR 是一个非常繁荣的研究与应用领域，几乎每天都有新技术报道。随着该领域逐步发展，研究者与开发者所面临的问题也相应地发生了变化。大部分的前期研究主要关心如何利用自动导出的图像特征从一个大图像集中检索出图像；目前的研究工作则集中在如何改进 CBIR 技术上，以研制出能满足用户实际需求的 CBIR 系统，研究的主要问题包括新类型的图像特征、表示方法及相似性度量等。显然，CBIR 系统不可能只用传统的文字查询方式达到，但这并不意味着不能将文字查询包含在

CBIR 系统中。事实上,已经有一些研究工作将基于文本的检索与 CBIR 集成在一起,即将图像的文字描述与图像的低层特征集成在一个 CBIR 系统中。事实上,研究人员已经发现全自动的特征提取似乎是不可能的,而使用者有时也无法明确地描述其中的一些图像特征,越来越多的研究工作已转向“半自动”的特征提取及交互式查询方式。现有的交互式 CBIR 系统广泛采用了相关反馈技术及机器学习技术。另外,一些研究者逐渐意识到“图像相似”是一个与人的视觉感知有关的概念,然而现有的 CBIR 技术无法完整地建立图像的低层特征与语义信息之间的联系,因此就提出了“语义图像检索”技术。目前针对该技术的研究也取得了大量的研究成果。考察 CBIR 技术的发展过程不难发现,CBIR 技术的发展已经历了几个发展阶段,每个阶段均体现了人们对“图像相似”概念理解的变化,并且均得到相应技术的支持(包括数据库技术、信息检索、计算机视觉、模式识别等)。

目前,CBIR 研究主要集中在以下几个方面。

### 1. 图像低层特征提取和描述技术

最初的图像检索研究主要集中在如何选择合适的图像全局特征去描述图像内容和采用什么样的图像度量方法进行图像匹配。尽管目前仍然无法有效地从图像低层特征获取高层语义,但是,如何提取一些有效的图像特征描述子以便于检索和存储却有着实际的研究价值。MPEG-7 标准中的视觉特征描述子部分(vvisual descriptor)就是专门针对图像低层特征(如颜色、纹理和形状)进行研究的,目前许多研究团体和研究机构也正从事这方面的研究。

### 2. 基于区域的图像检索

基于区域的图像检索的主要思想是通过图像分割技术(包括自动及半自动方法)将图像划分为不同的区域,然后对于每一个区域使用局部特征来进行描述,综合区域的局部特征从而得到图像的总特征描述,最后使用合适的相似性度量标准来检索图像。由于该方法更加贴近于用户查询时的思路,同时可以看作是由图像低层特征到图像语义特征的一个过渡,因此成为目前图像检索技术的一个重要研究方向。但由于自动图像分割是一个相当困难的技术,目前还无法使分割出的区域与图像中的对象很好地对应起来,因此目前借助于图像分割技术进行基于区域的图像检索的检索准确率并不高。但随着图像分割技术的发展,以及通过人工参与的半自动的分割方式,基于区域的图像检索无疑是一种非常有效的图像检索方法。同时本书第4章介绍的视觉注意计算模型、第5章介绍的自动图像标注技术,均是在基于区域特征提取的基础上完成的。

### 3. 图像语义特征提取

基于内容的图像检索技术所指的内容主要包括图像的颜色、形状、纹理和语义等

特征。其中,图像的颜色、形状和纹理等特征具有相对直观的特点,而语义特征具有相对主观抽象的特点。就图像检索技术的本身来说,基于语义特征的图像检索方式是最合理的图像检索方式,也最符合用户的需求。

语义检索是“智能”CBIR 系统的主要标志,图像语义检索技术的研究与发展可能是 CBIR 走向成熟与实用的关键。图像语义检索技术的研究要大量借鉴相关领域的研究成果,如人工智能、信息检索、认知科学等学科领域,并期待研究出新的知识表示方法以建立图像的低层特征与高层语义(概念描述)之间的联系。人工智能在语义特征的表示方面已经有许多成果,可以应用在 CBIR 中分析图像视觉特征和图像语义特征之间的映射关系,使得高层概念和低层视觉特征之间沟通成为可能,以缩短人机之间对相似图像理解的差距。此外,目前国际标准化组织制定的 MPEG-7 标准,其目标就是实现集高层语义特征和低层视觉特征于一体的基于内容的多特征综合检索。在研究图像语义检索技术时,需要采用认知科学的研究成果来分析图像内容的特征和人对图像的认知。图像信息在人脑中的长期记忆为心像,人对心像的记忆、检索等操作过程实际上是形象思维过程,因此形象思维科学中关于心像的表征和计算模型将对 CBIR 提供一定的指导。

但目前这种检索方式还面临 3 方面的问题<sup>[9]</sup>:①必须提供图像语义的有效描述方式;②必须有提取图像语义描述的方法;③语义检索系统要有合适的语义处理方法。因此,虽然人们更偏爱语义查询,但是这种查询方式的完全智能化目前还较难实现。目前,基于语义特征的图像检索主要研究如何从多种渠道获取图像语义信息,所获取的语义信息如何与图像低层特征结合并通过相关反馈在图像之间传递语义信息,以及如何将图像低层特征与图像的关键词结合进行图像的自动标注以提高检索的准确率等。

#### 4. 高维索引技术

网络的飞速发展导致产生了大型的图像数据库,但最新的研究模型也只能处理几百或几千幅图像,因为只有这样,在顺序扫描处理这些图像时才不至于严重影响系统的操作性能。多维索引方法在特征维数较低的情况下具有很好的检索性能,但在维数足够高的情况下,检索性能下降很快,其效率甚至会低于最原始的顺序查找方法。这一现象已经在几乎所有的传统索引方法中得到印证。随着图像数量的日益增多,检索速度已经成为瓶颈。尽管在这一研究领域已取得一些进展,但探索更加有效的高维索引技术仍是一个急需解决的问题。本书的第 6 章从子空间特征提取方面探讨了降维技术。

#### 5. 相关反馈技术

计算机视觉、模式识别与 CBIR 的基本区别就在于人在系统中的作用不同,前者依赖于计算机的处理能力企图达到“全自动”,而后者加入了人的协同工作,是一种“半自动”。在 CBIR 系统中,人是不可缺少的一个组成部分,因此需要探索人与机器

的协同工作。早期研究侧重于“完全自动化的系统”，并且试图寻找一种“最佳图像特征”，然而这样的思路并没有获得成功的 CBIR 系统。越来越多的研究工作则侧重于交互式系统，并且强调人在系统中的作用。现有的交互式 CBIR 系统主要采用“相关反馈”(relevance feedback)技术。相关反馈技术主要基于人机交互的思想，借助一种相关反馈的技术来猜测用户的需求，并且根据用户的需求动态调整系统检索时所采用的特征向量或参与检索的不同特征的权重系数，从而尽量缩小低层特征和高层语义之间的差距，提高算法的检索效果。其实，相关反馈是文本检索领域中一个基本的技术，Rui Yong 最先将其用到 CBIR 领域，实验证明它十分有效。近年来，将反馈技术用于图像检索中也是图像检索技术研究的一大热点，许多反馈方法被提了出来并用于图像检索中。目前，相关反馈技术中的主要的问题包括人机交互方式、用户模型、相关判断、学习算法等。

## 6. 相关反馈与机器学习结合技术

由于图像检索系统的最终用户是人，因此通过交互手段来捕获人对图像内容的认知是相当重要的。为了把用户模型嵌入图像检索系统中，最近几年在基于内容的图像检索领域引入了相关反馈与机器学习(machine learning)机制，将成熟的学习算法与图像检索中的在线学习过程(on-line learning)结合起来以提高检索准确率。代表性的工作包括基于 Bayesian 理论、基于 SVM(Support Vector Machine)、基于 Active Learning 等方法。

## 7. 性能评价

目前，基于内容的图像检索技术的性能评价主要借鉴了文本检索中的一些评价方法。由于图像内容的主观性使得对图像检索效果的评价仍没有通用的测试图像集和公认的评价标准。另外，因为缺少在相同数据集和查询条件下对不同检索系统进行有效性的对比实验，也难以确定什么样的评价标准更为有效，因此检索性能的评价还处在很不成熟的阶段，需要进一步研究。

对检索系统的合理评价，需要有一套能够平衡表达各种场景和事物的标准测试数据来评价检索的效率和效果，就像在图像处理领域，大家都用 Lena 图像作为实验图像一样。当然，性能评价是一项复杂的工作，要召集该领域专家及收集大量有代表意义的图像数据，以便能够测试各种算法的效率，并在此基础上定义标准的性能的评价准则，这样才可以利用标准的检索性能评价准则来全面地评价检索算法的性能。

## 1.2 CBIR 的研究内容

目前,有关 CBIR 的研究内容很多,但归纳起来,主要集中在 3 个方面:图像特征提取与匹配、索引机制及用户接口。

### 1.2.1 特征提取与匹配

---

所谓特征提取,就是从图像中把那些图像自身的内容信息提取出来,使用户可以据此进行图像检索。特征提取是 CBIR 研究的核心内容,大量的研究工作都是围绕这个主题开展的。

图像的特征主要包括低层特征和语义特征。低层特征主要包括图像的颜色、形状、纹理和空间关系等一些定量的特征,这些特征可以通过计算机自动或人机交互的方法来提取。对于所提取特征的高维特性,还需要通过有效的降维方法进行降维处理。语义特征是一种定性特征,是对图像内容的抽象描述,语义特征主要通过人工或人机交互(如相关反馈、机器学习等方法)的方法提取。

在提取完图像特征后,图像检索的主要任务就变成度量图像特征间的相似度问题。合理的相似性度量方法也是执行有效图像检索的关键。常用的相似性度量方法主要包括欧氏距离、城区距离、二次式距离、直方图相交法等。不同的相似性度量方法也有不同的优点和缺点,并非某一种度量方法对所有的图像检索系统均适用,它们也有各自的适用范围。图像检索时,应根据所提取特征的特点选择合适的相似性度量方法<sup>[3]</sup>。

### 1.2.2 索引机制

---

索引技术是加速图像相似性检索的关键技术之一,也是多媒体和数据库领域的研究热点和难点。针对多维数据索引机制,研究者提出了许多有效的方法。由于多种因素的影响(如数据的类型、分布情况、维数的高低等),那些适用所有场合、所有数据分布情况的索引结构是不存在的<sup>[4]</sup>。

为了有效完成图像检索,还必须解决图像特征的存储格式问题。对于图像的特征,它们之间可能没有内在的顺序,也可能具有多重相关特征,因此在图像检索系统中选用合适的数据结构模型对于有效的图像检索是十分必要的。目前常用的数据结构模型有 R 树、R<sup>+</sup>树、R<sup>\*</sup>树、X 树、SS 树、SR 树、k-d-B 树、四叉树、哈希法等。在这些数据结构模型中,每一种数据结构模型都有其内在的优点和缺点,因此在选择合适的

数据结构模型时需要考虑所提取的图像特征的特性。

上述索引方法在图像特征维数较低的情况下具有很好的检索性能，但后来的研究却发现<sup>[4]</sup>，这些索引方法在维数足够高的情况下，检索性能下降很快，会退化到顺序查找方法，导致“维数灾难”现象。当图像集的规模变得越来越大时，检索速度就会成为检索系统中的性能瓶颈，这时就需要采用高维数据索引方法来加速检索过程。

### 1.2.3 用户接口

图像检索系统的最终结果是交给用户鉴别的，因此在图像检索系统中，用户接口也起着重要作用，它在用户和检索系统之间提供了一种交互式的检索机制。用户可以通过该接口选取合适的查询机制及浏览图像检索结果，检索系统也可以根据用户的反馈结果进行学习，进一步提高系统的检索性能。一个理想的视觉查询的描述方式，不仅要方便于用户使用，精确地体现出用户的意图，同时又要包含足够的易于计算机接受的描述性信息。

用户接口提供的常用检索方式主要包括以下几种<sup>[5]</sup>。

(1) 草图查询 (query by sketching): 系统提供一个可以画草图的窗口，用户将想要查找的图像以草图的形式画出来并染上相应的颜色，系统从中抽取特征并进行检索。这种方式能提供给用户更大的想象和发挥空间，IBM 公司研制的图像和动态景象检索系统 QBIC 就提供了这种查询方式。

(2) 示例查询 (query by example): 这是 CBIR 中最常用的查询方式，就是由用户提供检索示例图像，如通过扫描仪输入图像或由用户从系统中选择查询图像，查询系统根据示例图像自动提取其特征，然后在图像库中找出与示例图像相似的图像并反馈给用户。该方式为用户提供了一种简便的方式来表达图像的内容，目前 Google 及 Baidu 图片搜索引擎均支持该方式。

(3) 类别浏览 (browsing by categories): 当用户对要查找的图像比较含糊或用户不熟悉需要查询图像的具体内容时，可以先按系统的分类体系浏览图像库，待发现感兴趣的目标后再进行示例查询。

(4) 图像特征和权值选取 (feature selection and weighting): 用户根据自身的检索要求，选择图像特征及不同特征的重要性 (权值) 作为检索依据。用户可以选择单个的也可以是复合的特征，同时各种特征还可以附加不同的权值。例如，“查找蓝色占 50%、红色占 50% 的图片”。这种方式在 IBM 的 QBIC 系统里得到了较好的体现。

(5) 相关反馈 (relevance feedback) 与机器学习 (machine learning): 用户根据需求，首先对系统的检索结果按照系统的提示进行标识 (如由用户标识检索结果中相似或不相似的图像)，系统根据用户标识信息，通过系统反馈及学习，进一步返回满足用户要求的图像。

## 1.3 CBIR 的关键技术

### 1.3.1 基本检索原理

在 CBIR 中，图像内容的描述借助于从图像中自动提取的视觉特征，检索过程是视觉特征间的相似匹配。假设在基于特征的图像相似性查询中，图像经过特征提取转换成为一个  $n$  维向量空间内的点，这样一个图像数据库就是一个  $n$  维数据空间内的点集，同时，数据库是完全动态的，允许进行动态的插入和删除。设数据库内对象的个数为  $m$ ，则数据库可表示为  $\{p_1, \dots, p_i, \dots, p_m\}$ ，其中  $p_i$  表示数据库中第  $i$  个对象的特征向量。根据应用领域的不同，多维数据库的查询方式也各不相同。对于给定的数据库，基于内容的相似性检索可以划分为 3 种类型<sup>[6]</sup>。

#### 1. 范围查询 (range query)

给定查询点  $q$ ，查询距离门限为  $t$ ，根据距离度量方式  $d$  (距离度量函数)，范围查询将查询出所有与点  $q$  距离小于等于  $t$  的点  $p$ 。

$$\text{RangeQuery}(q, t) = \{p \in \text{DB} \mid d(q, p) \leq t\} \quad (1-1)$$

其中，DB 表示图像数据库。基于范围的查询方式，其结果集的大小不能事先确定。一个用户在定义了查询距离门限  $t$  后，并不能确定可得到多少查询结果。这样的话有可能走向两种极端结果：不能得到任何结果或者得到数据库中的大多数点。

#### 2. $k$ 近邻查询 ( $k$ nearest-neighbor query)

$k$  近邻查询 (又称为  $k$  最近邻查询) 是从数据库中选择  $k$  个距离查询点最近的点作为结果集，可以形式化地描述如下。

$$\begin{aligned} \text{kNN}(q, k) = \{p_1, \dots, p_k \in \text{DB} \mid \forall p' \in \text{DB} \setminus \{p_1, \dots, p_k\}, \\ d(q, p_i) < d(q, p'), 1 \leq i \leq k\} \end{aligned} \quad (1-2)$$

在多媒体数据库中的信息检索应用中， $k$  近邻查询是最典型的应用。

#### 3. 限定误差范围内查找 ( $\alpha$ -cut 查找)

限定误差范围内查找是指从数据库中找出与检索样本的相似度不小于  $\alpha$  的所有图像。

通常范围查询需要比较复杂的检索界面，而且一般要求用户有关于图像特征的专



业知识与背景，所以这种检索多见于专业图像检索系统（如医学图像检索及诊断系统）。 $k$  近邻查询和 $\alpha$ -cut 查找是图像检索系统中最常见的检索模式，通常采用简单的QBE（Query by Example）检索界面。

### 1.3.2 图像内容及检索层次

图像检索的结果是为了满足用户的需求，这就需要所提取的图像内容符合人类对图像的认知特点。根据认知心理学研究，人类的认知过程具有层次性和整体性的特点，而不是简单的特征组合。因此，对图像内容与检索来说，也具有一定的层次性。

#### 1. 图像内容层次

对图像数据来说，所谓的内容具有多个层次上的含义<sup>[7]</sup>。

（1）感知层（perceptive level）：感知层特征往往指视觉上图像的颜色、纹理、形状、轮廓等特征，这些特征属于图像的低层特征。

（2）认知层（cognitive level）：认知层特征主要指图像中的主体、对象及对象间的关系等。认知层特征的提取往往首先需要通过图像分割，获取图像中的不同对象，然后提取不同对象的特征及对象间的关系。

（3）情感层（affective level）：情感层特征主要是指个人对图像内容的理解，并往往包含个人的情感、观点及意图等因素，如印象、情绪、感情等。

黄祥林等将图像内容理解为一个简化了的层次模型<sup>[8]</sup>，如图 1.4 所示。其中，第一层是图像的物理层特征，如颜色、纹理、轮廓和形状等；第二层是逻辑语义层特征，反映了图像所描述对象的标识及其空间关系等；第三层是抽象语义层特征，是人们对图像内容在认知层的概括和描述。

上述图像内容层次的划分方法实际上具有共同特点。另外，学者们往往把第一种划分方法中的认知层和情感层及第二种划分方法中的第二层和第三层概括为“语义层次”，而把第一层和“语义层次”之间的距离称作图像检索的“语义鸿沟”。图像语义的另一个重要特征是它的面向用户的特性，不同知识背景的人有不同的语义需求，并且对同一幅图像有时会产生截然不同的理解。如何在检索系统中体现这种差别，是图像语义表示的一个重要问题。目前，有关图像的语义特征，主要包括以下几点。

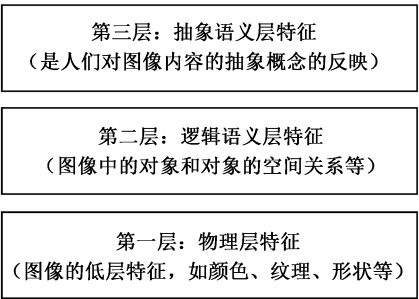


图 1.4 图像内容的层次模型

- (1) 情感语义：由图像带来的人的感觉，如兴奋、平静、高兴等。
- (2) 行为语义：由图像所表达的行为，如 NBA 球赛、世界杯足球赛等。
- (3) 场景语义：图像所处的场景，如日出、日落、下雨、下雪等。
- (4) 空间关系语义：图像中两个或多个对象之间的关系，如 A 在 B 的前面、汽车在房子的后面等。
- (5) 对象语义：图像中的对象，如花朵、运动员、房屋、山脉等。

## 2. 图像检索层次

根据图像内容的层次模型，用户的检索需求也分为 3 个层次<sup>[9]</sup>。

(1) 第一层次：利用图像的颜色、纹理和形状等低层特征及其组合，如颜色直方图、对比度、亮度等特征，通过图像的视觉相似性来进行检索。从本质上说，这个层次的特征并不能非常明显地区分不同情感。

(2) 第二层次：根据图像的逻辑特征信息，进行一定的逻辑推理和识别出图像中包含的对象类别。这个层次的检索需求可以是检索一个既定类型的物体，如“找一张航天飞机的图片”，也可以是检索一个独一无二的人或物，如“找一张自由女神像的图片”。要让计算机识别某一类的对象，首先必须让计算机认识该类对象，即获得对象概念的计算机内部表示，然后找出图像中可能是对象的区域，再来判定对象的类别。对于对象间的空间位置等关系，则是在识别出了对象的基础上来描述它们之间的拓扑关系。

(3) 第三层次：根据图像的抽象特征构成检索模式，包括物体或场景的描述及由此推理出来的场景语义、行为语义和情感语义。这个层次的检索需求可以是检索被命名的事件或活动，如“查找公园里玩耍的孩子”，也可以是检索具有情绪特点的图像，如“查找一张描述高兴的图片”。要回答这一类检索需求，就需要复杂的推理和主观判断，需要抽象地描述图像内容及复杂的情感认知计算。这类推理和判断往往建立在知识和学习的基础之上，常常要利用心理学和认知科学方面的一系列成果。

这种对用户查询进行层次分类的方法对于描述不同检索技术的能力及其局限性大有帮助。3 个层次最主要的差别体现在第一层次和第二层次之间，即是否真正利用了图像的语义。许多研究者将第二层次和第三层次的图像检索称为语义图像检索。

---

### 1.3.3 常用特征描述方法

对于图像检索中常用的低层视觉特征，本小节进行了简单介绍，详细内容在第 2 章中介绍。

## 1. 颜色特征

颜色特征是在图像检索中应用最为广泛的视觉特征，主要原因在于颜色往往和图像中所包含的物体或场景十分相关。另外，与其他视觉特征相比，颜色特征对图像本身的尺寸、方向、视角的依赖性较小，从而具有较强的鲁棒性。因此，基于颜色的图像表示方法自然就成为一种主要的图像索引技术，并得到相当广泛和深入的研究。

### 1) 全局颜色特征

在基于内容的图像检索中，应用最广泛的颜色特征是颜色直方图<sup>[10]</sup>。直方图容易计算，并且具有图像内容的旋转和平移不变性。同时，各种针对颜色直方图的改进方法也很多，如主色直方图<sup>[11]</sup>、累加直方图<sup>[12]</sup>、模糊直方图<sup>[13]</sup>、颜色差异直方图<sup>[14]</sup>等。另一种非常简单且有效的颜色特征描述方法是颜色矩（color moments）<sup>[12]</sup>。这种方法的数学基础在于图像中任何颜色分布均可以用它的矩来表示，由于极低的特征维数使其具有很强的竞争力。根据颜色直方图特性和信息论中信息熵的概念，John<sup>[15]</sup>提出采用图像颜色熵来表示图像的颜色特征；针对颜色熵和颜色矩的缺点，我们也提出了相应的改进办法<sup>[16]</sup>。

### 2) 空间颜色特征

上述颜色描述符所描述的特征是图像的全局颜色特征，不包括图像颜色的空间分布特征，因此仅仅利用这些特征进行图像检索极易造成误检现象。为此，多种空间颜色描述符被提了出来，如颜色聚合向量（color coherence vectors）<sup>[17]</sup>、颜色相关图（color correlograms）<sup>[18]</sup>、颜色几何分布直方图<sup>[19]</sup>、颜色分布熵<sup>[20]</sup>、感兴趣点颜色特征<sup>[21]</sup>、位平面颜色特征<sup>[22]</sup>等。结合空间结构关系的另一种方法是将图像划分为预先指定的多个区域，通过比较各个区域的直方图来提高空间判别能力。目前从划分局部区域的角度来说，常用的划分方法包括基于固定块的图像分割、基于手工的区域分割、采用交互半自动的区域分割及一些自动的颜色分割方法。同时，局部区域颜色信息主要采用平均颜色、主颜色、颜色直方图、颜色矩和二进制颜色集等来表示<sup>[23-25]</sup>。

## 2. 纹理特征

纹理蕴含着丰富的视觉信息，尽管还没有一种确切的定义，但纹理能很容易地被人所感知，并且在机器视觉领域得到了广泛的研究。而且人们发现，纹理特征具有不依赖于颜色或亮度的反映图像中同质现象的视觉特征，可以从微观上区分图像中不同的物体。常用的纹理描述方法有统计法、频谱法、结构法和模型法。

### 1) 统计法

统计法分析纹理的主要思想是通过图像中灰度级分布的随机属性来描述纹理特征。常用的方法有灰度共生矩阵<sup>[26]</sup>、对比度 (contrast)、粗糙度 (coarseness)、方向性 (directionality)、线像度 (line likeness)、规整度 (regularity) 和粗略度 (roughness)<sup>[27]</sup>。Penatti 等分析了 24 种颜色和 28 种纹理描述符在图像检索中的应用<sup>[28]</sup>。近年来, 一种简单有效的纹理描述方法——局部二值模式 (Local Binary Patter, LBP) 得到了广泛研究和应用<sup>[29, 30]</sup>。

### 2) 频谱法

频谱法主要借助于频率特性来描述纹理特征, 将空间域的纹理图像变换到频率域中, 利用信号处理的方法提取纹理特征。常用的频谱法主要包括傅里叶功率谱法<sup>[5]</sup>、Gabor 变换<sup>[31]</sup>、Wavelet<sup>[32]</sup>等。

### 3) 结构法

结构法分析纹理的基本思想是假定纹理模式由纹理基元以一定的、有规律的形式重复排列组合而成, 因此特征提取就变为确定这些基元并定量地分析它们的排列规则<sup>[33]</sup>。结构法分析的好处是纹理构成容易理解, 适合于高层检索, 描述规则的人工纹理。但对不规则的自然纹理, 由于基元本身提取困难及基元之间的排布规则复杂, 因此结构法受到很大的限制。

### 4) 模型法

模型法主要有随机场方法和分形法两种, 这些模型的共同特点是通过少量的参数表征纹理。常见的随机场模型有 Gauss-Markov 模型、Gibbs 模型、自回归纹理模型 (simultaneous auto-regressive) 等<sup>[34]</sup>。

## 3. 形状特征

形状是图像的重要可视化内容, 是人类视觉系统进行物体识别时所需要的关键信息之一。它不随周围环境 (如亮度等) 的变化而变化, 是物体的稳定信息。人们对一幅图像的理解很大程度上有赖于对图像中目标形状的区别和感知。在二维图像空间中, 形状通常被认为是一条封闭的轮廓曲线所包围的区域。因此, 形状特征包括对轮廓和区域的描述, 前者只用到形状的外边界, 而后者则关系到整个形状区域。从方法上来讲, 图像的形状视觉特征提取算法可以分为基于轮廓和基于区域两种。

### 1) 基于区域的描述方法

基于区域的形状描述方法是利用区域内的所有像素集合来获得描述目标轮廓所包围的区域性质的参数。常用的方法有几何不变矩、Legendre 矩、Zernike 矩、旋转矩、复数矩、通用傅里叶描述符等。Hu<sup>[35]</sup>基于形状不变矩提出了一系列分别具有变换、旋转和缩放无关性的 7 个矩。Teague<sup>[36]</sup>采用了更一般形式的 Legendre、Zernike、伪 Zernike 等正交多项式作为矩变换核代替传统的矩变换核。Zhang<sup>[37]</sup>提出了广义傅里叶描述符 (GFD) 的算法, 该算法对图像的像素在极坐标上进行 2D 傅里叶变换, 并使用变换后的傅里叶系数为特征向量。除此之外, MPEG-7 的 ART 描述子在极坐标的单位圆面上采用角半径变换 (Angular Radial Transformation, ART) 来获得矩不变量。除上述描述符外, 常用的基于区域的描述符还有区域的面积、圆度、欧拉数、离散度、偏心率、区域骨架等方法<sup>[38]</sup>。

### 2) 基于轮廓的描述方法

基于轮廓的形状描述方法是对包围目标区域的轮廓的描述。基于轮廓的图像特征一般可采用谱描述子、尺度空间滤波、多边形近似等一些方法。常用的方法有傅里叶描述符<sup>[37]</sup>、小波描述符<sup>[39]</sup>、曲率尺度空间描述符 (Curvature Scale Space Descriptor, CSSD)<sup>[40]</sup>。除上述轮廓描述方法外, 常用方法还有 Freeman 链码<sup>[41,42]</sup>、Sketch<sup>[43]</sup>、Shape Contexts<sup>[44]</sup>、Principal Curve<sup>[45]</sup>、Bag of Contour<sup>[46]</sup>、Triangel-Area Representation (TAR)<sup>[47]</sup>、边界矩<sup>[48]</sup>等。简单的边界几何形状不变量 (如周长、长轴、短轴、主轴方向等) 都可以用来描述形状的某一个特征, 且具有旋转不变等特性。

## 4. MPEG-7 中的图像特征描述符

在 MPEG-7 标准中考虑了 5 类基本的视觉特征, 对应地使用了 5 类描述符: 颜色描述符、形状描述符、纹理描述符、运动描述符和位置描述符。这里主要介绍颜色描述符、形状描述符及纹理描述符。

### 1) 颜色描述符

MPEG-7 中的颜色描述符有颜色空间描述符 (color space descriptor)、颜色量化描述符 (color quantization descriptor)、主颜色描述符 (dominant color descriptor)、可伸缩颜色描述符 (scalable color descriptor)、颜色布局描述符 (color layout descriptor)、颜色结构描述符 (color-structure descriptor) 及帧图/图组颜色描述符 (group of frames/group of pictures color descriptor) 等。其中, 颜色空间描述符和颜色量化描述符是两个辅助性的颜色描述符, 它们往往配合其他颜色描述符使用。

## 2) 形状描述符

MPEG-7 中定义的形状描述符有区域形状描述符 (region shape descriptor)、轮廓形状描述符 (contour shape descriptor) 及三维形状描述符 (shape 3D) 3 种。

区域形状描述符的表达式是由一系列 ART 系数构成的, ART 定义了一组二维的复值正交基函数, 将二维区域投射到这些基函数上, 得到的系数归一化后就可以描述区域的形状并用于匹配。它也是一种非常紧凑、有效的描述方式, 并具备分割噪声的功能。轮廓形状描述符是利用轮廓的曲率尺度空间 (Curvature Scale Space, CSS)<sup>[49]</sup> 来描述封闭的轮廓。三维形状描述符可用于相对自然的或虚拟的三维目标。在描述三维物体的形状特征时, 首先建立物体的三维网格 (3D mesh) 模型, 然后在物体表面的局部区域计算出该处的形状指数 (shape index)。

## 3) 纹理描述符

MPEG-7 中的纹理描述符包括同质纹理描述符 (homogenous texture descriptor)、纹理浏览描述符 (texture browsing descriptor) 和边缘直方图描述符 (edge histogram descriptor) 3 种。

同质纹理描述符通过在频域计算能量和能量方差来提供对纹理的量化描述, 它采用 5 个尺度和 6 个方向的 30 个 Gabor 滤波器对纹理图像进行多分辨率分解, 将频域内滤波器组输出能量的均值和标准差作为纹理特征。纹理浏览描述符从类似于人类感知的角度对纹理的方向性 (directionality)、规则性 (regularity) 和粗糙程度 (coarseness) 进行描述, 适用于图像的浏览和根据纹理粗糙程度进行的分类。边缘直方图描述符描述边缘的空间分布信息, 首先将图像划分成 16 个互不重叠的矩形区域, 对每个图像区域分别按水平、垂直、45° 角、135° 角 4 个方向和 1 个无方向性边缘 5 类信息进行直方图统计。此描述符具有尺度不变性, 支持纹理旋转和旋转不变匹配, 适用于非一致纹理图像。

## 5. 语义特征

研究表明, 人类视觉系统对图像的理解和识别, 除了基于图像低层特征之外, 更重要的是其所表达的场景本身的结构和层次关系。实际上, 用人类的智能来识别一个目标, 并非仅仅建立在目标的低层视觉特征 (如颜色、形状和纹理) 上, 而是充分考虑了目标所描述的对象、事件, 甚至情感等语义。要较好地满足用户对目标进行识别或检索的需求, 需要在目标特征的描述上充分考虑高层语义<sup>[50,51]</sup>。目前典型的方法有 Bag of Words<sup>[52]</sup>、Semantic Tags<sup>[53]</sup>、自动图像标注<sup>[54]</sup>、Sparse CCA (Sparse Canonical Correlation Analysis)<sup>[55]</sup>、融合文本和视觉特征<sup>[56]</sup>等。

图像语义特征的提取与理解是弥补语义鸿沟的有效途径, 虽然相关研究在本领域取得了一定的研究进展, 但图像语义理解仍是目前一个极富挑战性的研究问题。

### 1.3.4 特征匹配技术

图像检索的匹配策略大致可以分为两种，一种是完全匹配，另外一种是相似性匹配。当两幅图像的特征完全相同时，图像匹配成功，称之为完全匹配。当两幅图像的特征间的距离小于某一个阈值时，图像匹配成功，称之为相似性匹配。在基于内容的图像检索中占主导地位的是建立在图像低层视觉特征对比基础上的相似性检索。在提取图像特征后，可采用相应的相似性度量策略来进行特征匹配，也就是通过确定待检索图像同数据库目标图像特征向量间的距离来确定待检索图像同数据库中目标图像间的相似性。一个合适的相似性度量方法对图像检索结果影响很大。相似性度量方法的好坏会影响到图像检索的性能，相似性度量的计算复杂度会影响到图像检索的用户响应时间。理想的相似性度量方法应该满足人的视觉特性，也就是说视觉上相似的图像间应具有较小的距离，而视觉上不相似的图像间应具有较大的距离。

#### 1. 变量公理

设  $A, B, C$  为任意的  $n$  维特征向量，通常情况下，距离度量函数  $d$  应受以下 4 条公理的限制<sup>[5]</sup>。

##### 1) 自相似公理

$$d(A, A) = d(B, B) = 0 \quad (1-3)$$

##### 2) 最小公理

$$d(A, B) \geq d(A, A) = 0 \quad (1-4)$$

##### 3) 对称公理

$$d(A, B) = d(B, A) \quad (1-5)$$

##### 4) 三角不等公理

$$d(A, C) \leq d(A, B) + d(B, C) \quad (1-6)$$

在实际应用中，所采用的相似度比较函数并非严格满足上述距离度量的 4 条公理，它们往往只是满足上述公理的某个或某几个。

#### 2. 常用的匹配算法

目前，图像检索中用到的特征匹配算法很多，常用的有以下几种。

### 1) Minkowsky 距离

Minkowsky 距离是基于  $L_p$  范数定义的, 即

$$L_p(\mathbf{A}, \mathbf{B}) = \left( \sum_{i=1}^n |a_i - b_i|^p \right)^{\frac{1}{p}} \quad (1-7)$$

如果  $p=1$ ,  $L_1(\mathbf{A}, \mathbf{B})$  称为城区 (city-block) 距离, 即

$$L_1(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^n |a_i - b_i| \quad (1-8)$$

如果  $p=2$ ,  $L_2(\mathbf{A}, \mathbf{B})$  称为欧氏距离 (euclidean distance), 即

$$L_2(\mathbf{A}, \mathbf{B}) = \left[ \sum_{i=1}^n (a_i - b_i)^2 \right]^{\frac{1}{2}} \quad (1-9)$$

如果  $p \rightarrow \infty$ ,  $L_\infty(\mathbf{A}, \mathbf{B})$  称为 Chebychv 距离, 即

$$L_\infty(\mathbf{A}, \mathbf{B}) = \max_{i=1}^n |a_i - b_i| \quad (1-10)$$

### 2) 直方图相交法

直方图相交法 (histogram intersection) 是由 Swain 等人于 1991 年首次提出的。直方图相交法计算简单快速, 并且能较好地抑制背景的影响, 其数学描述为<sup>[10]</sup>

$$d(\mathbf{A}, \mathbf{B}) = 1 - \sum_{i=1}^n \min(a_i, b_i) \quad (1-11)$$

式 (1-11) 可以进一步进行归一化处理, 得

$$d(\mathbf{A}, \mathbf{B}) = 1 - \frac{\sum_{i=1}^n \min(a_i, b_i)}{\min(\sum_{i=1}^n a_i, \sum_{i=1}^n b_i)} \quad (1-12)$$

### 3) 二次式距离

对基于颜色直方图的图像检索来说, 二次式距离 (quadratic distance)<sup>[57]</sup>已被证明比欧氏距离及直方图相交法更为有效, 其原因在于这种距离考虑到了不同颜色之间存在的相似度。二次式距离可以表示为

$$d_{\text{qad}}(\mathbf{A}, \mathbf{B}) = (\mathbf{A} - \mathbf{B})^T \mathbf{M} (\mathbf{A} - \mathbf{B}) \quad (1-13)$$

其中,  $\mathbf{M} = [m_{ij}]$ ,  $m_{ij}$  表示直方图中下标为  $i$  和  $j$  的两种颜色之间的相似度。这种方法通过引入颜色相似性矩阵  $\mathbf{M}$ , 使其能够考虑到相似但不相同的颜色间的相似性因素, 颜色相似性矩阵  $\mathbf{M}$  可以通过对颜色心理学的研究获得。



## 4) 余弦距离 (cosine distance)

余弦距离计算的是两个向量间方向的差异, 其定义为

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{A}^T \mathbf{B} = |\mathbf{A}| \cdot |\mathbf{B}| \cos \theta \quad (1-14)$$

$$d_{\cos}(\mathbf{A}, \mathbf{B}) = 1 - \cos \theta = 1 - \frac{\mathbf{A}^T \mathbf{B}}{|\mathbf{A}| \cdot |\mathbf{B}|} \quad (1-15)$$

其中,  $|\mathbf{A}| = (\mathbf{A}^T \mathbf{A})^{\frac{1}{2}}$ ;  $|\mathbf{B}| = (\mathbf{B}^T \mathbf{B})^{\frac{1}{2}}$ 。

## 5) 相关系数

相关系数是一个可以用来表征两个向量之间线性关系紧密程度的量, 定义为

$$\rho(\mathbf{A}, \mathbf{B}) = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}} \quad (1-16)$$

其中,  $\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i$ ;  $\bar{b} = \frac{1}{n} \sum_{i=1}^n b_i$ 。

采用相关系数, 两个向量间的距离可表示为

$$d_{\rho} = 1 - \rho(\mathbf{A}, \mathbf{B}) \quad (1-17)$$

## 6) Kullback-Leibler 散度和 Jeffrey 散度

Kullback-Leibler (K-L) 散度定义为

$$d_{kl}(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^n a_i \lg \frac{a_i}{b_i} \quad (1-18)$$

其中,  $a_i \geq 0$ ,  $b_i \geq 0$ , 且  $\sum_{i=1}^n a_i = 1$ ,  $\sum_{i=1}^n b_i = 1$ 。K-L 散度的缺点是非对称并对直方图柱值数敏感。改进后的 Jeffrey 散度具有对称性和对噪音及直方图柱值数的健壮性, 其定义为<sup>[58]</sup>

$$d_{jef}(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^n \left( a_i \lg \frac{a_i}{m_i} + b_i \lg \frac{b_i}{m_i} \right) \quad (1-19)$$

其中,  $m_i = \frac{a_i + b_i}{2}$ 。

7)  $\chi^2$  距离

$\chi^2$  距离的定义为

$$d_{\chi^2}(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^n \frac{(a_i - m_i)^2}{m_i} \quad (1-20)$$

其中,  $m_i = \frac{a_i + b_i}{2}$ 。

#### 8) EMD (Earth Mover's Distance) [59]

形象地解释 EMD 就是: 空间  $S$  中分布着  $m$  堆土  $\mathbf{p}_i$  ( $i=1, \dots, m$ ), 每堆土的质量为  $w_{\mathbf{p}_i}$ , 同时分布有  $n$  个土坑, 土坑的大小为  $\mathbf{q}_j$  ( $j=1, \dots, n$ ), 每个土坑可以装土的质量为  $w_{\mathbf{q}_j}$ , 即  $\mathbf{p} = \{(\mathbf{p}_1, w_{\mathbf{p}_1}), \dots, (\mathbf{p}_m, w_{\mathbf{p}_m})\}$ ,  $\mathbf{q} = \{(\mathbf{q}_1, w_{\mathbf{q}_1}), \dots, (\mathbf{q}_n, w_{\mathbf{q}_n})\}$ 。把所有的土填到这些坑内, 做的功可表示为

$$\text{work}(\mathbf{p}, \mathbf{q}, f) = \sum_{i=1}^m \sum_{j=1}^n d(\mathbf{p}_i, \mathbf{q}_j) f_{ij} \quad (1-21)$$

其中,  $d(\mathbf{p}_i, \mathbf{q}_j)$  表示第  $i$  堆土到第  $j$  个坑的距离, 为了体现图像的距离测度的区别, 称之为基本距离;  $f_{ij}$  表示第  $i$  堆土运到第  $j$  个坑的土的质量;  $d(\mathbf{p}_i, \mathbf{q}_j) f_{ij}$  表示把第  $i$  堆土中质量为  $f_{ij}$  的土运到第  $j$  个坑所做的功。上式隐含的约束条件包括:

- (1)  $f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n$ ;
- (2)  $\sum_{j=1}^n f_{ij} \leq w_{\mathbf{p}_i}, 1 \leq i \leq m$ ;
- (3)  $\sum_{i=1}^m f_{ij} \leq w_{\mathbf{q}_j}, 1 \leq j \leq n$ ;
- (4)  $\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min(\sum_{i=1}^m w_{\mathbf{p}_i}, \sum_{j=1}^n w_{\mathbf{q}_j})$ 。

条件 (1) 说明每次搬运土的质量大于零时才做功; 条件 (2) 说明从第  $i$  堆土运到各个坑的土的质量, 一定不会大于该堆土的质量; 条件 (3) 说明第  $j$  个坑能够接受的土的质量, 一定不会大于该坑能够接受的土的最大量。从而, EMD 定义为

$$\text{EMD}(\mathbf{p}, \mathbf{q}) = \frac{\sum_{i=1}^m \sum_{j=1}^n d(\mathbf{p}_i, \mathbf{q}_j) f_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}} \quad (1-22)$$

从上述定义可以看出, 我们不难把图像相似性度量的问题, 转化为计算 EMD 的问题。当测量两幅图像的距离时, 可以把一幅图像的特征向量映射为土堆, 把另一副图像的特征向量映射为土坑, 则两幅图像间的距离, 就是把所有土填入土坑内在选择最佳路径的条件下做功的最小值。

## 9) 编辑距离 (edit distance)

编辑距离又称为 Levenshtein 距离, 被定义为从一个串变换到另一个串的最少插入、删除、替换操作。设  $A$  表示有限的字符集合,  $\varepsilon$  表示一个空符号。编辑操作可以表示为  $a \rightarrow b, a \rightarrow \varepsilon, \varepsilon \rightarrow a$ , 其中  $a, b \in A$ ,  $a \rightarrow b$  代表替换操作,  $a \rightarrow \varepsilon$  代表删除操作,  $\varepsilon \rightarrow a$  代表插入操作。设  $c(a \rightarrow b), c(a \rightarrow \varepsilon), c(\varepsilon \rightarrow a)$  分别表示各类操作的代价, 设符号  $x = x_1 x_2 \cdots x_n$ ,  $y = y_1 y_2 \cdots y_m$ , 则它们间的编辑距离的计算可表示为

步骤 1  $D(0, 0) = 0$

步骤 2 For  $j = 1, \cdots, m$

$$D(0, j) = D(0, j-1) + c(\varepsilon \rightarrow y_j)$$

步骤 3 For  $i = 1, \cdots, n$

$$D(i, 0) = D(i-1, 0) + c(x_i \rightarrow \varepsilon)$$

步骤 4 For  $i = 1, \cdots, n$ , For  $j = 1, \cdots, m$

$$D(i, j) = \min \left\{ \begin{array}{l} D(i-1, j-1) + c(x_i \rightarrow y_j) \\ D(i-1, j) + c(x_i \rightarrow \varepsilon) \\ D(i, j-1) + c(\varepsilon \rightarrow y_j) \end{array} \right\}$$

从上述的计算步骤可以看出, 串  $x$  和串  $y$  间的编辑距离  $d(x, y) = D(n, m)$ 。

## 10) Hausdorff 距离

Hausdorff 距离是一种定义于两个点集上的最大-最小 (max-min) 距离, 它主要用于测量两个点集的匹配程度。给定两个有限点集  $P = \{p_1, p_2, \cdots, p_m\}$ ,  $Q = \{q_1, q_2, \cdots, q_n\}$ , 则  $P$  与  $Q$  之间的 Hausdorff 距离定义为

$$D(P, Q) = \max \{d(P, Q), d(Q, P)\} \quad (1-23)$$

其中,  $d(P, Q)$  为从点集  $P$  到点集  $Q$  的有向 Hausdorff 距离,  $d(Q, P)$  可进行类推。

$$d(P, Q) = \max_{p \in P} \min_{q \in Q} \|p - q\| \quad (1-24)$$

$$d(Q, P) = \max_{q \in Q} \min_{p \in P} \|q - p\| \quad (1-25)$$

式中,  $\|\cdot\|$  为定义在点集合上的某种距离范数。

这里, 式 (1-23) 称为双向 Hausdorff 距离, 是 Hausdorff 距离的最基本形式; 式 (1-24) 及式 (1-25) 分别称为从  $P$  集合到  $Q$  集合和从  $Q$  集合到  $P$  集合的单向 Hausdorff 距离, 双向 Hausdorff 距离是单向距离  $d(P, Q)$  和  $d(Q, P)$  两者中的较大者, 它度量了两个点集间的最大不匹配程度。

在上述的相似性度量方法中, 没有任何一种方法可以适用于所有特征向量间的相似性度量, 其主要原因是上述度量方法具有特征依赖的特点, 不同的特征应该应用不

同的度量方法。例如，直方图相交法不适合于非直方图的特征；虽然二次式距离可以有效地度量颜色直方图的距离，但对其他特征向量的距离度量效果却没有采用欧氏距离度量的效果好。因此，在进行图像检索时，显然难以证明上述相似性度量方法谁更具优势，在具体的应用过程中只有根据实际需要进行灵活选择。需要指出的是，在相似性度量的能力上，某些已知工作给出了一些结果。例如，文献[37]指出，除了颜色直方图特征外，在进行特征向量间的相似性度量时，欧氏距离和城区距离相对于其他相似性度量方法具有更好的检索性能。

### 3. 精确查询与近似查询

精确相似性查询是指在给定特征向量和距离度量函数的前提下，获得精确的查询结果；而近似相似性查询是指在一定的误差允许下，获得相似性查询结果。通过控制查询结果的不确切性，采用近似相似性查询可以较高地提升查询性能。目前，近似近邻查询技术是高维数据检索的一个新的研究趋势，它提供了一种克服维数灾难现象的新的手段和方法。在高维情况下，绝大多数的索引结构对于大型多媒体数据库，其查询性能都很难达到预期效果。采用近似近邻查询，可以使用户在查询结果的精确性和查询时间上取得折中。但是，目前对于近似查询并没有严格的定义和统一的评价标准。研究者提出了许多近似查询的索引结构和搜索算法<sup>[4]</sup>，在这其中，具有严格定义的是 $\varepsilon$ 近似近邻查询。

给定查询向量 $\mathbf{q}$ 和一个近似系数 $\varepsilon (\varepsilon > 0)$ ，如果满足下式，则称向量 $\mathbf{p}$ 是查询向量 $\mathbf{q}$ 的 $(1+\varepsilon)$ 近似近邻。

$$d(\mathbf{p}, \mathbf{q}) \leq (1+\varepsilon)d(\mathbf{p}^*, \mathbf{q}) \quad (1-26)$$

其中， $\mathbf{p}^*$ 是查询向量 $\mathbf{q}$ 的真正近邻。也就是说， $\mathbf{p}$ 与真正近邻的相对误差为 $\varepsilon$ 。更进一步，对于第 $k$ 个 $(1+\varepsilon)$ 近似近邻与真正 $k$ 近邻的相对误差也是 $\varepsilon$ ，即

$$d(\mathbf{p}^k, \mathbf{q}) \leq (1+\varepsilon)d(\mathbf{p}^{k*}, \mathbf{q}) \quad (1-27)$$

其中， $\mathbf{p}^k$ 和 $\mathbf{p}^{k*}$ 分别表示 $\mathbf{q}$ 的第 $k$ 个近似近邻和真正近邻。

在多媒体信息检索应用中，特征向量和距离测度的选取都带有主观性或者试探性，特征向量本身就只是多媒体内容的近似表示，而且并不能从数学上对对象之间的相似性进行严格的定义。由于多媒体信息所蕴含内容的丰富性和用户需求判断的主观性，很难找到一种精确的图像内容表示和度量方法。所以，精确近邻搜索并不具有精确含义。在许多查询应用领域，都采用了人机交互式的查询方式，搜索引擎需要根据用户对查询的反馈信息重新调整查询过程。比较典型的应用是基于相关反馈的多媒体检索系统。在此类系统中，查询结果允许有一定的不精确性，但是查询的响应时间是至关重要的。对用户来讲，可以容许一定的查询误差，来获得更好的查询性能。

### 1.3.5 稀疏表示技术

1959年 David Hubel 和 Torsten Wiesel<sup>[60]</sup>通过对猫的视觉皮层简单细胞感受野的研究认为视觉皮层细胞感受野采用稀疏表示的原则表示视觉感知信息。在此基础上, Olshausen 和 Field<sup>[61]</sup>就人脑视觉信息这一问题进行了深入研究, 指出人类视觉系统捕获自然场景中的信息只需特定几个视觉神经元即可, 也就是说人类视觉系统能够利用较少的神经元表示所观察到的客观事物, 从而为图像的稀疏表示奠定了基础。

稀疏表示的基本思想是认为自然图像或者信号可以由一个基函数字典线性叠加表示<sup>[62]</sup>。给定信号  $\mathbf{Y} \in \mathbf{R}^n$ , 基函数字典  $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m\} \in \mathbf{R}^{n \times m}$ , 则该信号可用下述稀疏模型表示, 即

$$\min \|\mathbf{X}\|_0 \quad \text{s.t.} \quad \mathbf{Y} = \mathbf{DX} \quad (1-28)$$

其中,  $\mathbf{X}$  为信号  $\mathbf{Y}$  在字典  $\mathbf{D}$  上的稀疏表示或稀疏编码;  $\|\mathbf{X}\|_0$  是  $L_0$  范数, 表示  $\mathbf{X}$  中非零分量的个数。

对于式 (1-28), 当字典  $\mathbf{D}$  的维数满足  $n > m$  时, 线性方程  $\mathbf{Y} = \mathbf{DX}$  是超定方程,  $\mathbf{X}$  有唯一解。当  $n < m$  (字典  $\mathbf{D}$  是过完备字典) 时, 线性方程  $\mathbf{Y} = \mathbf{DX}$  是欠定方程,  $\mathbf{X}$  的解是不唯一的, 此时需要添加相应的约束条件, 如式 (1-29) 和式 (1-30) 所示。在此情况下, 式 (1-28) 的求解问题变为了对式 (1-29) 或式 (1-30) 的优化问题。

$$\min \|\mathbf{Y} - \mathbf{DX}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{X}\|_0 \leq k \quad (1-29)$$

$$\min \|\mathbf{X}\|_0 \quad \text{s.t.} \quad \|\mathbf{Y} - \mathbf{DX}\|_2 \leq \varepsilon \quad (1-30)$$

可以看出, 式 (1-29) 是采用不超过  $k$  个原子的线性组合表示的条件下的稀疏模型, 并使得逼近误差尽可能小。而式 (1-30) 是在  $\mathbf{Y}$  满足逼近误差的条件下的稀疏模型 ( $\varepsilon$  是逼近精度)。

为了实现信号的稀疏表示, 关键要实现对上述稀疏表示模型的求解。为此需要解决两个基本问题: 一个是稀疏分解, 即稀疏系数的求解; 另一个是稀疏字典学习, 即如何构造过完备字典使得信号的表示是稀疏的。

#### 1. 稀疏分解

稀疏分解的目的是为了得到最优的信号稀疏表示。Mallat 和 Zhang<sup>[63]</sup>首次提出信号在过完备字典上分解的思想, 并指出信号在过完备字典上的分解结果一定是稀疏的, 这为信号的稀疏分解奠定了基础。目前常用的稀疏分解算法主要分为两类: 一类是近似逼近算法, 即贪婪算法; 另一类是将目标函数进行转化, 从而简化求解过程。

### 1) 贪婪算法

贪婪算法的基本原理是通过迭代方式选择字典中与信号最为相似的原子来近似逼近原始信号,常用的方法有匹配追踪 (Matching Pursuit, MP)、正交匹配追踪 (Orthogonal Matching Pursuit, OMP)<sup>[64]</sup>等。MP 算法在每一次迭代过程中,从完备原子库里选择与信号最为相似的原子来表示信号,并求出残余量,然后选出与残余量最为匹配的原子。该算法在有限维条件下满足指数级收敛,但是推广到无限维,该算法并不收敛。另外,MP 算法并不对已选出的全部原子进行正交化,而仅与最近选出的原子正交,因此在已选字典原子构成的子空间上信号每次迭代的结果可能是次最优的。也就是说,MP 算法不能保证信号的最优逼近。而 OMP 算法在每次迭代时通过递归对已选择的全部原子进行正交化以保证迭代的最优性。但是,OMP 算法并不能实现对所有信号的精确重构,为此,Neeldell 等在 OMP 算法的基础上提出了正则化正交匹配 (Regularized Orthogonal Matching Pursuit, ROMP) 算法<sup>[64]</sup>,该算法采用约束等距性以保证解的稳定性,从而实现对满足该条件的原子集与稀疏信号进行精确重构。与 OMP 算法相比,ROMP 算法首先选择多个原子作为候选集,然后从候选集中按照正则化原则挑选出部分原子,最后将其并入最终的原子集,从而实现原子的快速、有效的选择。随后出现的压缩采样匹配追踪 (Compressive Sampling Matching Pursuit, CoSaMP) 算法<sup>[65]</sup>和子空间追踪 (Subspace Pursuit, SP) 算法<sup>[66]</sup>引入了回溯思想用以降低重建复杂度,这些算法的重建质量与线性规划方法相当。上述分解算法的前提是已知信号的稀疏度,但实际应用中稀疏度往往是未知的。为此,Do 等人提出了稀疏自适应匹配追踪 (Sparsity Adaptive Matching Pursuit, SAMP) 算法<sup>[67]</sup>,该算法通过设置一个可变步长逐步逼近进行重建,从而可以在稀疏度未知的情况下获得较好的重建效果,速度也快于 OMP 算法。

### 2) 目标函数转化

由于直接根据稀疏模型求解信号的稀疏表示已被证明是 NP-hard 难题,为此许多研究者期望通过目标函数转化方法,将对 0-范数的求解问题转化为其他形式的函数进行求解。目前常用的转化方法有  $L_p$  范数近似法算法<sup>[68,69]</sup>及  $NSL_0$  (Newton Smooth  $L_0$  Norm)<sup>[70]</sup>算法等。

## 2. 字典构造

构造合适的字典在稀疏表示中具有重要作用,目前字典构造方法主要分为两类:一类是基于数学模型的字典构造方法,另一类基于自适应的字典构造方法。

### 1) 基于数学模型的字典构造

基于数学模型的字典构造方法关键是选择合适的生成函数。目前,常用的生成函

数主要包括脊波变换、Curvelets 变换、Contourlets 变换、Bandelets 变换等<sup>[71,72]</sup>。

## 2) 基于自适应的字典构造

基于自适应的字典构造方法主要通过对样本的学习实现。这类方法虽然缺乏理论指导,计算量大,但同基于数学模型的字典构造方法相比,该类方法自适应能力强,实验效果明显,因此基于自适应的字典构造方法成为目前使用最广泛的方法。常用的自适应字典构造方法有最大似然法(maximum likelihood methods)<sup>[73]</sup>、最优方向法(Method of Optimal Directions, MOD)<sup>[74]</sup>、最大后验概率法(Maximum A-Posteriori Probability Approach, MAP)<sup>[75]</sup>及 K-SVD 算法<sup>[76]</sup>等。

### 1.3.6 性能评价准则

在进行图像检索时往往需要选择一种或多种最有效的特征描述方法和相似性度量方法,这就需要对不同的图像特征或特征组合及不同的相似性度量方法的检索效果进行全面的评价,比较不同方法的性能,找出最好的方法。但是,由于图像检索具有很强的主观性,同时也很难找到一个统一的图像测试库,因此评价一个图像检索算法性能的优劣并不容易。这里列举了目前图像检索领域广泛应用的几个公认的图像检索算法评价准则。

#### 1. 精确度和检索率

精确度(precision)和检索率(recall)是目前在 CBIR 中应用最为广泛的一种评价准则。精确度的含义是在一次查询过程中,系统返回的相关图像数目占所有返回图像数目的比例。如果在检索结果集合中,正确相关图像数目多,则精确度就高。检索率则指系统返回的查询结果中相关图像数目占图像库中所有相关图像数目(包括返回的和没有返回的)的比例。

设  $S$  为图像库中所有和查询图像相关的图像集,  $R$  为所有检索到的图像集合,  $s$  为一次查询中检索到的所有相关图像数目,  $u$  为一次检索过程中检索到的不相关的图像数目,  $v$  为图像库中和检索图像相关但在检索中未被检索到的图像数目,这样精确度和检索率可表示为

$$\text{recall} = P(R | S) = \frac{P(S \cap R)}{P(S)} = \frac{s}{s + v} \quad (1-31)$$

$$\text{precision} = P(S | R) = \frac{P(S \cap R)}{P(R)} = \frac{s}{s + u} \quad (1-32)$$

精确度和检索率越高,表明该检索系统的效果越好。一般,检索率和精确度是一对矛盾,当要求精确度较高时,检索率较低,反之亦然。因此,一般的检索系统只要

求在这两者之间达到一个最优的平衡点, 就认为达到了较好的检索性能。

另外, 对精确度和检索率来说, 首先需要知道数据库中每一类图像的相似图像数目, 因此对于大型图像库, 尤其对图像库中图像动态变化的图像库来说, 要做到这一点比较困难。另外, 要统计检索率和精确度还需要用户在检索结果中标记出与示例图像相似的图像, 由于用户的主观性, 对同一次查询来说, 不同用户得到的精确度和检索率可能并不相同。

## 2. 命中准确率

精确度和检索率需要用户在图像库中人工找出与查询图像相似的图像集, 这将耗费大量的人工劳动, 因此这种度量准则对于较小型的图像数据库比较合适。如果图像库测试集已经提前进行了分类, 如 Corel Image Gallery 等类型的数据库, 就可以简单地将每一个图像类别作为其中每一幅图像的相关图像, 由此来度量算法的检索准确率。设图像  $q$  所在的相关图像集为  $G$ , 图像检索算法自动输出了  $T$  幅相似图像, 其中命中  $G$  的有  $n$  幅图像, 此次检索的准确率定义为

$$P_T = \frac{n}{T} \quad (1-33)$$

由此, 平均多个查询的检索准确率就可以度量算法的检索性能。

## 3. 排序值评测法

设  $q$  是一幅查询图像,  $g_1, g_2, \dots, g_n$  为图像检索算法检索到的与  $q$  相关的且从主观上认为相似的图像, 设  $\text{rank}(g_i) (i=1, 2, \dots, n)$  为图像  $g_i$  在检索结果图像序列中对应的排序值, 则下述两个指标可以有效地衡量算法的检索性能。

$$\text{r-measure} = \frac{1}{n} \sum_{i=1}^n \text{rank}(g_i) \quad (1-34)$$

$$\text{p-measure} = \frac{1}{n} \sum_{i=1}^n \frac{i}{\text{rank}(g_i)} \quad (1-35)$$

其中, 第一个指标定义了所有相关图像在检索结果中的平均排序, 显然, 此指标越小, 检索算法的准确率越高; 第二个指标定义了所有相关图像在靠前列的紧密程度, 因此该值越大表明检索结果越好, 如果所有相关图像都排在最前面, 则此指标取值为 1。

## 4. ANMRR

ANMRR (Average Normalized Modified Retrieval Rank) 是 MPEG-7 推荐的一种性能评价方法。设  $N(q_i) (i=1, 2, \dots, Q)$  表示图像库中与图像  $q_i$  相似的所有图像数目,  $M = \max\{N(q_1), N(q_2), \dots, N(q_Q)\}$ ,  $K = \min\{4N(q_i), 2M\}$ , 设与例子图像相似的图像在检索结果序列中所处的位置为



$$\text{rank}(k) = \begin{cases} k, & k \leq K \\ K+1, & k > K \end{cases} \quad (1-36)$$

从而, ANMRR 定义为

$$\text{ANMRR} = \frac{1}{Q} \sum_{i=1}^Q \frac{\sum_{k=1}^{N(q_i)} \frac{\text{rank}(k)}{N(q_i)} - 0.5 - 0.5 \cdot N(q_i)}{K + 0.5 - 0.5 \cdot N(q_i)} \quad (1-37)$$

由上式可知, ANMRR 的取值越小表明该算法的检索性能越好。

### 5. 前 $N$ 个结果的正确率与检索率

前  $N$  个结果的正确率描述如下: 设  $\mathbf{R}$  为某一具有特定语义含义的图像集合, 设  $q_i \in \mathbf{R}$  为任一检索示例图像, 在一次检索过程中, 若系统返回  $N$  个结果 (记为  $I_1, I_2, \dots, I_j, \dots, I_N$ ), 则正确率  $P_N(q_i)$  定义为

$$P_N(q_i) = \sum_{j=1}^N \frac{\phi(I_j, \mathbf{R})}{N} \quad (1-38)$$

其中,  $\phi(I_j, \mathbf{R}) = \begin{cases} 0, & \text{if } I_j \notin \mathbf{R} \\ 1, & \text{if } I_j \in \mathbf{R} \end{cases}$ , 那么对于所有测试样例检索图像集得到的平均正确率可表示为

$$P_N = \sum_{i=1}^K \frac{P_N(q_i)}{K} \quad (1-39)$$

其中,  $K$  表示测试样例检索图像集中的图像数目。该正确率的定义简单地说就是在返回的前  $N$  个结果中有多高的比例是正确的。

前  $N$  个结果的检索率定义为

$$R_N(q_i) = \sum_{j=1}^N \frac{\phi(I_j, \mathbf{R})}{\|\mathbf{R}\|} \quad (1-40)$$

其中,  $\|\mathbf{R}\|$  表示图像集  $\mathbf{R}$  中的图像数目, 那么对于所有测试样例检索图像集得到的平均检索率可表示为

$$R_N = \sum_{i=1}^K \frac{R_N(q_i)}{K} \quad (1-41)$$

## 1.4 CBIR 的应用与经典系统

### 1.4.1 CBIR 的应用

---

CBIR 技术将对大规模图像信息的管理和访问提供有力支持。它可以广泛应用于信息检索服务、犯罪预防、医疗诊断、新闻和广告、商标和知识产权、地理信息和远程遥感、教育培训和军事等领域，目前比较成熟的应用有指纹识别、人脸识别和图像搜索引擎等。

#### 1. 信息检索

面对日益增长的图像信息和图像检索需求，传统的检索手段显得越发笨拙和不合时宜，而 CBIR 技术可以自动对图像库进行基于内容的索引，查找出与检索图像相似的图像，更好地满足基于内容的检索需求。

#### 2. 知识产权

科技的飞速发展使得人们越来越关注知识产权的保护问题，而许多知识产权的载体都是图像。例如，一个新商标在注册前需要和已经注册的商标进行比较以确定它是否与其他商标雷同，以免造成侵权行为，这就需要计算机对庞大的商标图像库进行视觉相似性检索，直接比较其颜色或形状特征来确定是否相似。

#### 3. 犯罪预防

通过指纹、鞋印、人脸来确认和查找凶犯是目前安全部门常用的手段。计算机检索系统根据指纹、鞋印和人脸的图案信息进行基于特征的相似性鉴别，这种技术已被世界各地安全部门广泛采用。

#### 4. 医疗诊断

现代医疗器械和技术的发展产生了大量的医用图像信息，如 X 光片等。人们发现通过以往相似病例的诊断图像来进行辅助诊断是一项非常有意义的事情，这就需要在图像库里查找出与当前诊断图像相似的图像。目前，医学图像检索已成为基于内容的图像检索技术应用的一个主要方向。

#### 5. 教育培训

在教育培训领域，如远程教学、交互式培训、自学教育及雇员再教育等，有着广

阔的应用前景。国外在教育培训领域已投入了大量的经费，开展了相关课题的研究工作。我国多媒体教学研究工作的开展，网上教学与辅导已进入实用阶段，这些都为图像数据库应用于教育培训领域提供了广阔的前景。

## 6. 数字图书馆

随着现代信息技术革命的深入发展，图书馆正在发生前所未有的变革，正在向数字图书馆迈进。数字图书馆实际上是一个数字信息资源库，其中有字符数值库、文本库、声音库、图像库等。因此，如何快速、高效地从数字图书馆中找出用户所需的信息就成为现代图书馆研究的热点和关键技术之一。

## 7. 工业与商业领域

工业应用包括企业多媒体信息系统、CAD/CAM 等，商业应用有电子商务、在线广告、在线购物、股票等。

## 8. 军事领域

CBIR 技术在军事领域有着巨大的应用需求，如从雷达图像里识别敌机、从卫星照片里识别目标、为巡航导弹提供制导系统等。此外，CBIR 技术在犯罪预防和信息检索领域里的应用也可以应用到军事领域里。

总之，CBIR 技术是一项快速发展的颇具发展潜力的前瞻性技术，在许多领域都具有很高的应用价值。

### 1.4.2 经典 CBIR 系统介绍

---

鉴于基于内容的图像检索系统的重要性、有效性和优越性，从 20 世纪 90 年代初期，各大公司和科研机构陆续推出了一些商用或研究用的图像检索系统。本节将简单介绍几个经典的软件系统。

#### 1. QBIC 系统

QBIC (Query by Image Content) 系统是由 IBM 公司于 20 世纪 90 年代开发制作的图像和动态景象检索系统，是在基于内容的图像检索领域应用最早的商用产品，它的系统框架和结构对后来的图像检索系统具有深远的影响。

QBIC 系统提供了多种查询方式，包括利用标准范图（示例图像）检索、用户绘制简图或扫描输入图像进行检索、选择色彩或结构查询方式、用户输入动态影像片段和前景中运动的对象检索。QBIC 系统中使用的颜色特征有色彩百分比、色彩位置分

布等；使用的纹理特征是根据 Tamura 提出的纹理表示的一种改进，即结合了粗糙度、对比度和方向性的特性；使用的形状特征有面积、圆形成度、偏心率、主轴偏向和一组代数矩不变量。另外，QBIC 系统还考虑到了高维特征的索引，采用  $R^*$  树作为索引结构。另外，QBIC 系统还支持基于文本的关键字查询方式。总之，该系统技术成熟，功能全面，为基于内容的图像检索技术的验证和推广作出了很大贡献，其检索示例和查询界面分别如图 1.5 和图 1.6 所示。

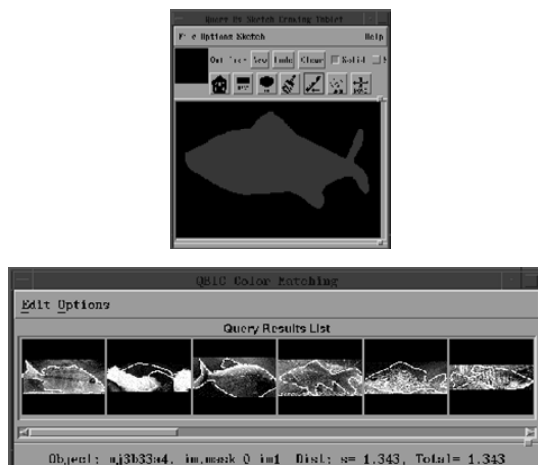


图 1.5 QBIC 系统基于形状特征的检索示例



图 1.6 QBIC 系统的查询界面

## 2. Photobook 系统

Photobook 系统<sup>[77]</sup>是 MIT 媒体实验室开发的图像检索系统。Photobook 系统有 3 个子部分，分别用于提取形状、纹理和面部特征。图像在装入时按人脸、形状或纹理特性自动分类，图像根据类别通过显著语义特征压缩编码。因此，用户可以在这 3 个子部分中分别进行基于形状、基于纹理和基于面部特征的图像检索。

Photobook 系统提供了一些交互式工具进行基于内容的浏览和检索，它使用颜色、纹理、形状、变换域系数建立特征向量，并提供多种相似性准则，如欧几里得距离、城区距离、向量内积、小波树距离等。用户也可以选择一种或几种算法的线性组合计算图像的相似性，而且还允许用户勾画感兴趣的区域，参与特征提取过程。Photobook 系统的查询界面如图 1.7 所示。

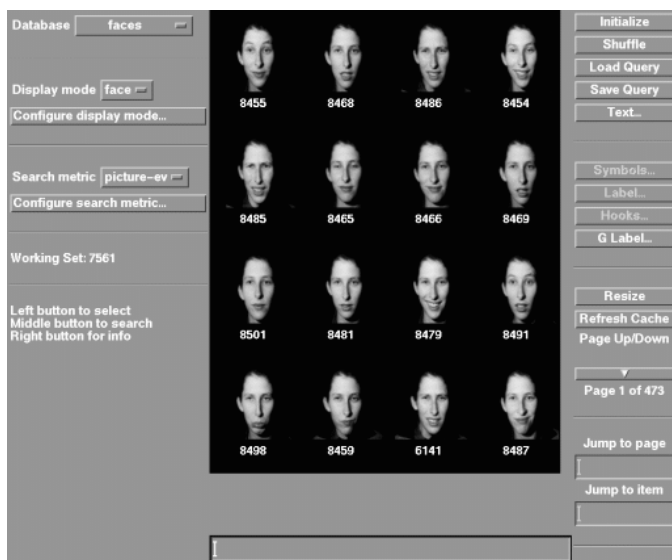


图 1.7 Photobook 系统的查询界面

## 3. VisualSEEK 系统和 WebSEEK 系统

VisualSEEK 系统和 WebSEEK 系统<sup>[78]</sup>是由美国哥伦比亚大学开发的姊妹系统，其主要特点是研究利用图像区域空间关系进行查询和从压缩域提取视觉特征来进行检索，系统中主要使用的特征是颜色特征和基于小波变换的纹理特征。

VisualSEEK 系统支持基于视觉特征和它们之间空间关系的查询，其查询界面如图 1.8 所示。WebSEEK 系统主要是面向 Web 查询的，采用了先进的特征提取技术，用户界面强大，操作简单，查询途径丰富，输出画面生动且支持用户直接下载信息，其查询界面如图 1.9 所示。

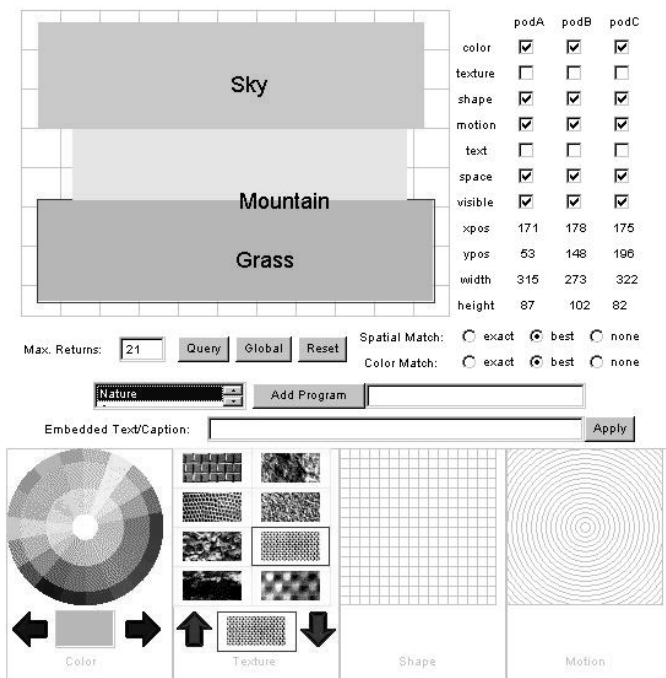


图 1.8 VisualSEEK 系统的查询界面

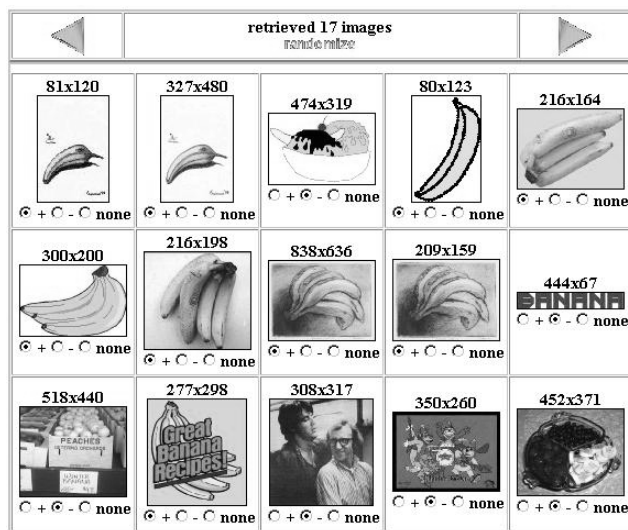


图 1.9 WebSEEK 系统的查询界面

#### 4. Blobworld 系统

Blobworld 系统<sup>[79]</sup>是由 UC Berkeley 大学开发的基于区域的图像检索系统。该系统使用的图像特征为颜色、纹理、位置及区域和背景的形状。在颜色方面使用 Lab 空

间的 218bin 的颜色直方图进行描述, 纹理通过区域的平均对比度和各向异性来描述, 区域形状的描述由面积、离心率和方向性组成。对各个特征的相似性分别采用欧式距和加权欧式距进行度量, 最后给出统一的相似性距离。同时, 该系统还对 218bin 的颜色直方图的计算进行降维处理, 并采用  $R^*$  树索引结构以提高检索效率。Blobworld 系统的查询界面如图 1.10 所示。

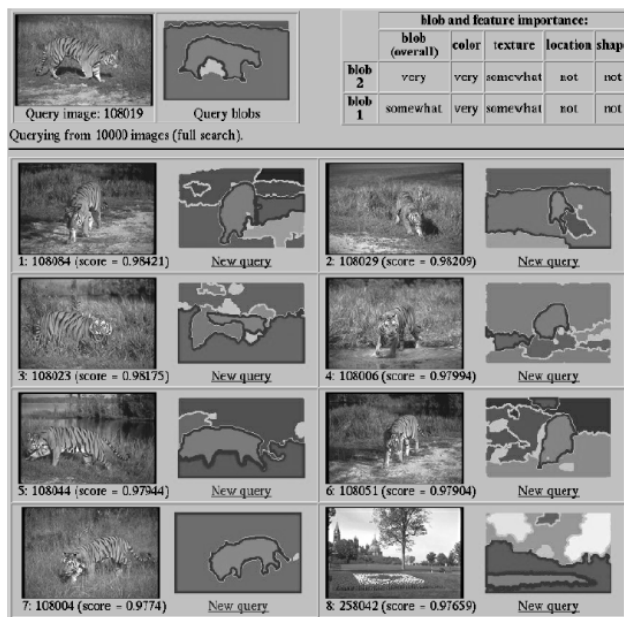


图 1.10 Blobworld 系统的查询界面

## 5. MARS 系统

MARS 系统<sup>[80]</sup>是由 UIUC (University of Illinois at Urbana-Champaign) 开发的支持图像低层特征的复合检索的图像检索系统。它的应用范围相当广泛, 包括计算机视觉、图像数据库检索和信息检索等多个领域。其特点是使用比较全面的图像低层特征, 提供基于树结构的多特征的组合检索。在图像特征方面, 使用 HSV 空间的 HS 上的色彩直方图来描述图像的颜色; 提取图像纹理的粗糙度和方向性及对比度等特征描述纹理; 采用图像的规则分割的方法对图像特征的空间分布进行描述 (颜色直方图和小波变换系数); 根据纹理对图像进行分割来实现图像中的对象描述, 并对分割后的对象区域按照敏感性进行分组; 使用 Fourier 描述子对图像中对象的形状进行描述。检索时, 对上述特征分别采用相应的相似性度量方法, 最终给出综合排名。由于采用多方面的图像特征描述与相似性度量方法, 该系统可以提供较复杂的检索功能, 如可以通过布尔表达式进行组合检索。MARS 系统的焦点不在于找到单一的最佳特征表达, 而是如何把不同的视觉特征组织成为一个可以动态适应不同应用和不同用户的有意义

的检索机制。这个系统的突出特点在于引入了相关反馈机制，能够根据与用户的交互，动态地组织和优化查询，提高检索效率，其检索效果如图 1.11 所示。



图 1.11 MARS 系统的检索效果

## 6. SIMPLIcity 系统

SIMPLIcity 系统<sup>[81]</sup>是由斯坦福大学与滨州大学开发的。系统提出了一种综合区域匹配 (Integrated Region Matching, IRM) 来定义图像的相似度。它具有旋转和平移不变性，允许区域间的一对多匹配，对不准确的分割比较稳健。图 1.12 给出了其检索界面，图 1.13 给出了其中的一次检索结果。

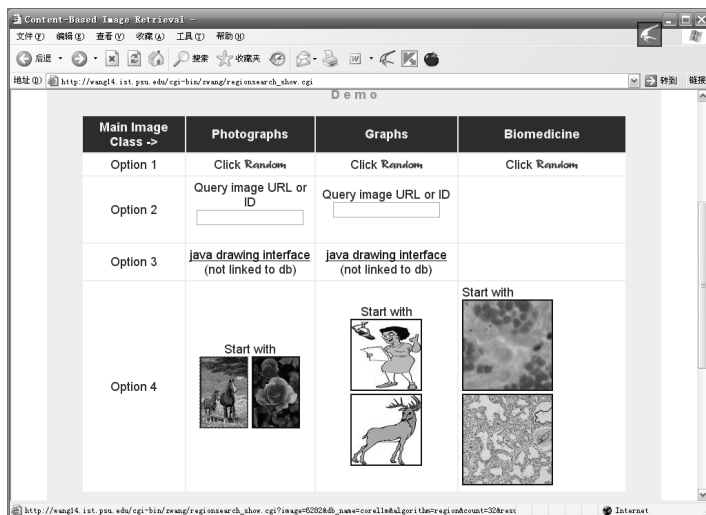


图 1.12 SIMPLIcity 系统的检索界面



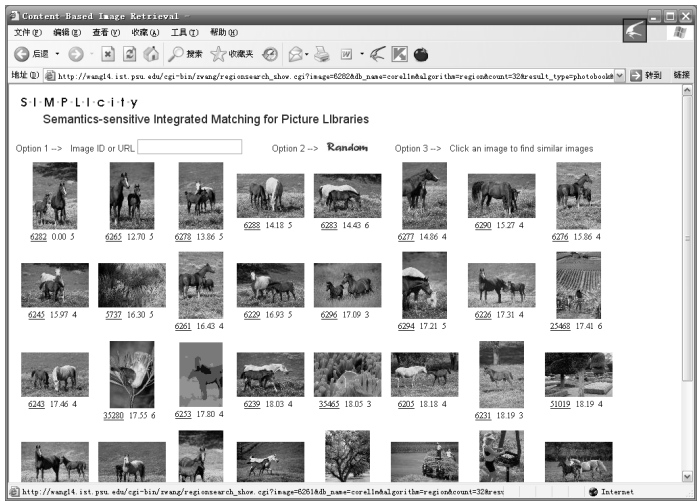


图 1.13 SIMPLicity 系统的一次检索结果

## 7. CIRES 系统

CIRES 系统<sup>[82]</sup>综合运用了图像的高层语义特征和低层视觉特征进行图像检索。语义特征主要运用了图像的结构信息，视觉特征主要采用了图像的颜色和纹理信息。CIRES 系统的检索示例如图 1.14 所示。

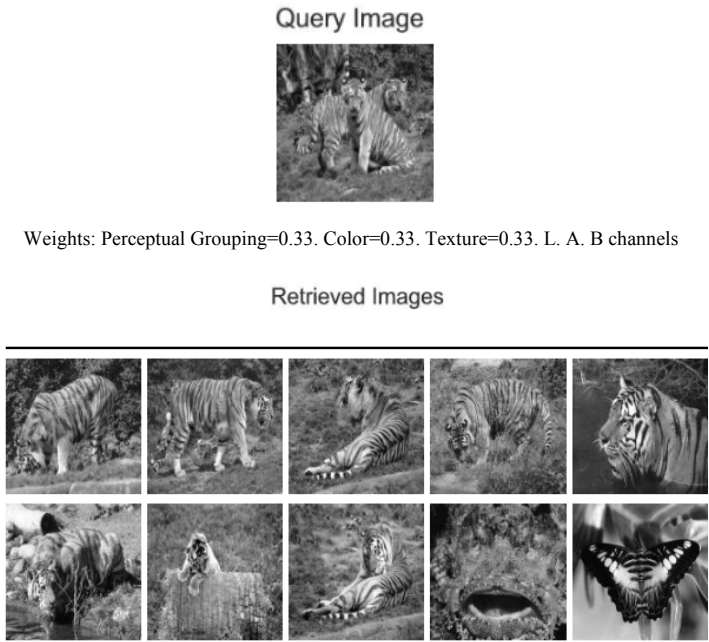


图 1.14 CIRES 系统的检索示例



图 1.14 CIRES 系统的检索示例 (续)

## 1.5 本书内容安排

国内外至今已经出版了大量有关基于内容的图像检索技术方面的文章,有关图像检索技术的综述文献也已发表了很多。国际上每年也都召开许多有关信息检索技术方面的学术会议,许多大会都有图像检索技术的主题和分会。早期的著作(章毓晋 2003, 庄越挺等 2002)对基于内容的图像检索技术作了一定的阐述,2007 年,清华大学出版社出版了一部针对基于内容的图像检索的专著(周明全等 2007),对基于内容的图像检索技术作了较全面的论述。我们在 2009 年的著作《图像低层特征提取与检索技术》中主要针对图像低层特征的提取与检索进行了全面的论述。但近年来,图像检索技术发展迅速,各种新技术及方法不断涌现。在此基础上,我们在总结传统图像检索技术的基础上,进一步对图像检索新技术进行了总结。

本书的内容主要安排如下。

第 1 章介绍了 CBIR 的发展与现状、CBIR 的研究内容与其所涉及的关键技术、CBIR 的应用及一些经典的 CBIR 系统。

第 2 章介绍了图像低层特征的提取与表达技术,主要涉及图像的颜色、形状和纹理 3 种基本特征,又介绍了 MPEG-7 中的图像特征描述符。

第 3 章介绍了基于压缩域的图像检索技术,包括空间压缩域和变换压缩域中常用的描述算法,并介绍了两种基于 DCT 压缩域的图像纹理及形状特征提取方法。

第 4 章介绍了视觉注意计算模型,引入了基于特征加权、基于高斯混合和基于 CIELab 的 3 种视觉注意计算模型。

第 5 章介绍了自动图像标注技术,主要讨论了图像视觉特征选择、低层特征到高层语义之间映射模型的建立两个方面的问题。

第 6 章针对图像检索中的维数灾难问题,详细讨论了子空间特征提取技术。

## 参 考 文 献

- [1] 孙君顶, 赵珊. 图像低层特征提取与检索技术[M]. 北京: 电子工业出版社, 2009.
- [2] 孙君顶. 基于内容的图像检索技术研究[D]. 西安: 西安电子科技大学, 2005.
- [3] 洪安祥. 基于内容的图像检索若干论题研究[D]. 杭州: 浙江大学, 2003.
- [4] 崔江涛. 高维索引技术中向量近似方法研究[D]. 西安: 西安电子科技大学, 2005.
- [5] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003.
- [6] 赵珊. 基于内容的图像检索关键技术研究[D]. 西安: 西安电子科技大学, 2007.
- [7] Hanjalic A. Video and image retrieval beyond the cognitive level: The needs and possibilities[C] // Photonics West 2001-Electronic Imaging. International Society for Optics and Photonics, 2001:130-140.
- [8] 黄祥林, 沈兰荪. 基于内容的图像检索技术研究[J]. 电子学报, 2002, 30(7):1065-1071.
- [9] 王惠锋, 孙正兴, 王箭. 语义图像检索研究进展[J]. 计算机研究与发展, 2002, 39(5):513-523.
- [10] Swain M J, Ballard D H. Color indexing[J]. Intl. J. on Computer Vision, 1991, 7(1):11-32.
- [11] Talib A, Mahmuddin M, Husni H, et al. A weighted dominant color descriptor for content-based image retrieval[J]. Journal of Visual Communication and Image Representation, 2013, 24(3):345-360.
- [12] Stricker M, Orengo M. Similarity of color images[C] // Proceedings of SPIE Storage and Retrieval for Image and Video Database, 1995, 2420:381-392.
- [13] Han J, Ma K K. Fuzzy color histogram and its use in color image retrieval[J]. IEEE Transactions on Image Processing, 2002, 11(8):944-952.
- [14] Liu G H, Yang J Y. Content-based image retrieval using color difference histogram[J]. Pattern Recognition, 2013, 46(1):188-198.
- [15] John Z M. An Information theoretic approach to content based image retrieval[D]. Baton Rouge: Louisiana State University and Agricultural and Mechanical College, 2000.
- [16] Sun J, Zhang X, Cui J, et al. Image retrieval based on color distribution entropy[J]. Pattern Recognition Letters, 2006, 27(10):1122-1126.
- [17] Pass G, Zabini R, Miller J. Comparing Images Using Color Coherence Vectors[C] // ACM International Conference on Multimedia, MA, 1996:65-73.
- [18] Huang J. Color-Spatial Image Indexing and Applications[D]. New York: Cornell

- University, 1998.
- [19] Lim S, Lu G. Spatial statistics for content based image retrieval[C] // International Conference on Information Technology: Coding and Computing, 2003:155-159.
  - [20] 孙君顶, 毋小省. 基于颜色分布特征的图像检索[J]. 光电子 • 激光, 2006, 17(8):1009 -1013.
  - [21] Wang X Y, Yang H Y, Li Y W, et al. Robust color image retrieval using visual interest point feature of significant bit-planes[J]. Digital Signal Processing, 2013, 23(4):1136 -1153.
  - [22] 孙君顶, 毋小省. 基于位平面熵及分布熵的图像检索[J]. 系统工程与电子技术, 2009, 31(3):719-722.
  - [23] 何清法, 李国杰. 综合分块主色和相关反馈技术的图像检索方法[J]. 计算机辅助设计与图形学学报, 2001, 13(10):912-917.
  - [24] Yoo H W, Jang D S, NA Y K. An efficient indexing structure and image representation for content-based image retrieval[J]. IEICE Transactions on INF&SYST, 2002: 1390-1398.
  - [25] Vimina E R, Jacob K P. Content Based Image Retrieval Using Low Level Features of Automatically Extracted Regions of Interest[J]. Journal of Image and Graphics, 2013, 1(1):7-11.
  - [26] Haralick R M, Shanmugam K. Texture features for image classification[J]. IEEE-SMC, 1973, 3(6):610-621.
  - [27] Tumara H, Mori S, Yamawaki T. Texture features corresponding to visual perception[J]. IEEE-SMC, 8(6):460-473.
  - [28] Penatti O A B, Valle E, Torres R S. Comparative study of global color and texture descriptors for web image retrieval[J]. Journal of Visual Communication and Image Representation, 2012, 23(2):359-380.
  - [29] Pietikäinen M. Computer vision using local binary patterns[J]. Springer, 2011.
  - [30] Brahmam S, Jain L C, Lumini A, et al. Local Binary Patterns: New Variants and Applications[J]. Springer Berlin Heidelberg, 2014.
  - [31] Manjunath B S, Ma W Y, Texture features for browsing and retrieval of image data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8):837-842.
  - [32] Choy S K, Tong C S. Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval[J]. IEEE Transactions on Image Processing, 2010,19(2):281-289.
  - [33] Young D C, Sang Y S, Nam C K. Image Retrieval using BDIP and BVLC Moments[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13 (9): 951-957.

- [34] C, Komodakis N, Paragios N. Markov Random Field modeling, inference & learning in computer vision & image understanding: A survey[J]. Computer Vision and Image Understanding, 2013, 117(11):1610-1627.
- [35] Hu M K. Visual pattern recognition by moment invariants[J]. IRE Trans. on Information Theory, 1962, 8(2):179-187.
- [36] Teague M R. Image analysis via the general theory of moments[J]. J. Opt. Soc. Am. 1980, 70 (8):920-930.
- [37] Zhang D S. Image Retrieval Based on Shape[D]. Melbourne: Monash University, 2002.
- [38] Zhang D, Lu G. Review of shape representation and description techniques[J]. Pattern recognition, 2004, 37(1):1-19.
- [39] 杨翔英, 章毓晋. 小波轮廓描述符及在图像查询中的应用[J]. 计算机学报, 1999, 22(7):752-757.
- [40] Mokhtarian F, Abbasi S, Kittler J, et al. Robust and efficient shape indexing through curvature scale space[C] // Proceedings of the British Machine Vision Conference, Edinburgh, UK, 53-62.
- [41] Iivariinen J, Visa A. Shape recognition of irregular objects[C] // Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, SPIE, 1996:25-32.
- [42] 孙君顶. 基于链码分布特征及相关性的轮廓描述与检索[J]. 光电子·激光, 2008, 19(8):1112-1115.
- [43] Shu X, Wu X J. A novel contour descriptor for 2D shape matching and its application to image retrieval[J]. Image and vision Computing, 2011, 29(4):286-294.
- [44] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4):509-522.
- [45] Ataer-Cansizoglu E, Bas E, Kalpathy-Cramer J, et al. Contour-based shape representation using principal curves[J]. Pattern Recognition, 2013, 46(4):1140-1150.
- [46] Wang X, Feng B, Bai X, et al. Bag of contour fragments for robust shape classification[J]. Pattern Recognition, 2014: 2116-2125.
- [47] Alajlan N, El Rube I, Kamel M S, et al. Shape retrieval using triangle-area representation and dynamic space warping[J]. Pattern Recognition, 2007, 40(7):1911-1920.
- [48] Sun J D, Zhang Z S. Shape retrieval based on combination moment invariants[C] // Proceedings of Information Technology and Environmental System Science, 2008, 3:301-305.

- [49] Farzin Mokhtarian, Alan K Mackworth. A Theory of Multiscale: Curvature-Based Shape Representation for Planar Curves[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(8):789-805.
- [50] Liu Y, Zhang D, Lu G, et al. A survey of content-based image retrieval with high-level semantics[J]. Pattern Recognition, 2007, 40(1):262-282.
- [51] Kaur H, Jyoti K. Survey of Techniques of High Level Semantic Based Image Retrieval[J]. IJRCCT, 2013, 2(1):015-019.
- [52] Wu L, Hoi S C H, Yu N. Semantics-preserving bag-of-words models and applications[J]. IEEE Transactions on Image Processing, 2010, 19(7):1908-1920.
- [53] Ma H, Zhu J, Lyu M R T, et al. Bridging the semantic gap between image contents and tags[J]. IEEE Transactions on Multimedia, 2010, 12(5):462-473.
- [54] Zhang D, Islam M M, Lu G. A review on automatic image annotation techniques[J]. Pattern Recognition, 2012, 45(1):346-362.
- [55] 庄凌, 庄越挺, 吴江琴, 等. 一种基于稀疏典型性相关分析的图像检索方法[J]. 软件学报, 2012, 23(5):1295-1304.
- [56] Chen Y, Sampathkumar H, Luo B, et al. iLike: bridging the semantic gap in vertical image search by integrating text and visual features[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(10): 2257-2270.
- [57] Hafner J, Sawhney H, Equitz W, et al. Efficient color histogram indexing for quadratic form distance functions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(7):729-736.
- [58] Puzicha J, Hufnann T, Buhmann J. Non-parametric similarity measure for unsupervised texture segmentation and image retrieval[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997:267-272.
- [59] Yossi R, Carlo T, Leonidas G. The Earth Mover's Distance as a Metric for Image Retrieval[J]. International Journal of Computer Vision, 2000, 40(2):99-121.
- [60] Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex[J]. The Journal of physiology, 1959, 148(3):574-591.
- [61] Olshausen B A, Field D J. Sparse coding with an overcomplete basis set: A strategy employed by V1[J]. Vision research, 1997, 37(23):3311-3325.
- [62] Olshausen B A. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. Nature, 1996, 381(6583): 607-609.
- [63] Mallat S G, Zhang Z. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12):3397-3415.

- [64] Needell D, Vershynin R. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit[J]. Foundations of computational mathematics, 2009, 9(3):317-334.
- [65] Needell D, Tropp J A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples[J]. Applied and Computational Harmonic Analysis, 2009, 26(3):301-321.
- [66] Dai W, Milenkovic O. Subspace pursuit for compressive sensing signal reconstruction[J]. IEEE Transactions on Information Theory, 2009, 55(5):2230-2249.
- [67] Do T T, Gan L, Nguyen N, et al. Sparsity adaptive matching pursuit algorithm for practical compressed sensing[C] // 2008 IEEE 42nd Asilomar Conference on Signals, Systems and Computers, 2008:581-587.
- [68] Chen S S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit [J]. SIAM review, 2001, 43(1):129-159.
- [69] 赵谦, 孟德宇, 徐宗本.  $L_{1/2}$  正则化 Logistic 回归[J]. 模式识别与人工智能, 2012, 25(5):721-728.
- [70] 赵瑞珍, 林婉娟, 李浩, 等. 基于光滑  $L_0$  范数和修正牛顿法的压缩感知重建算法 [J]. 计算机辅助设计与图形学学报, 2012, 24(4):478-484.
- [71] Do M N, Vetterli M. The contourlet transform: an efficient directional multiresolution image representation[J]. IEEE Transactions on Image Processing, 2005, 14(12): 2091-2106.
- [72] Pennec E L, Mallat S. Sparse geometric image representations with bandelets[J]. IEEE Transactions on Image Processing, 2005, 14(4):423-438.
- [73] Lewicki M S, Sejnowski T J. Learning overcomplete representations[J]. Neural computation, 2000, 12(2):337-365.
- [74] Engan K, Aase S O, Husoy J H. Method of optimal directions for frame design[C] // 1999 IEEE International Conference on Acoustics, Speech and Signal Processing, 1999, 5:2443-2446.
- [75] Kreutz-Delgado K, Murray J F, Rao B D, et al. Dictionary learning algorithms for sparse representation[J]. Neural computation, 2003, 15(2):349-396.
- [76] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11):4311-4322.
- [77] Pentland A, Picard R, Sclaroff S. Photobook: tools for content-based manipulation of image database[C] // Proceedings of SPIE, 1994, 2185:34-47.
- [78] Smith J R. Integrated spatial and feature image systems: retrieval, compression and analysis[D]. New York: Columbia University, 1997.

- [79] Cllad C, Serge B, Hayit G, et al. Blockworld: image segmentation using expectation-maximization and its application to image querying[J]. IEEE Transactions on PAMI, 2002, 24(8):1026-1038.
- [80] Rui Y, Huang T S, Mehrotra S. Relevance feedback: a Powerful tool in interactive content-based image retrieval[J]. IEEE Transactions on CSVT, 1998, 8(5):644-655.
- [81] James Z Wang, Jia L, Gio W. SIMPLicity: semantics-sensitive integrated matching for picture libraries[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(9):947-963.
- [82] Iqbal Q, Aggarwal J K. CIRES: A system for content-based retrieval in digital image libraries[C] // Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on. IEEE, 2002, 1:205-210.



## 图像低层特征的提取与表达

图像特征的提取与表达是 CBIR 技术的基础，获取有效的图像的低层特征是提取图像高层语义信息的关键环节。本章介绍了近年来常用的图像低层特征提取技术，主要包括颜色、形状和纹理 3 种特征。

### 2.1 颜色特征的提取与表达

颜色特征是在图像检索中应用最为广泛的视觉特征，也是人识别图像的主要感知特征，主要原因在于颜色往往和图像中所包含的物体或场景十分相关。自然界中，同一类物体通常有相同或相近的颜色特征，不同类的物体则可能表现为不同的颜色特征，因此颜色经常可以作为区分不同对象最为简单有效的一种手段。与其他的视觉特征相比，颜色特征对图像本身的尺寸、方向、视角的依赖性较小，从而具有较高的鲁棒性。因此，大多数图像检索系统都将颜色特征作为图像检索的主要手段。

针对颜色特征的提取与表达，首先，需要选择合适的颜色空间来描述颜色特征；其次，采用一定的量化方法将颜色特征表达为向量的形式；最后，以一定的方式来描述颜色特征。本节首先讨论了 CBIR 中常用的颜色空间及颜色量化的手段，然后在此基础上，介绍了目前常用的颜色特征描述方法。

#### 2.1.1 颜色空间

从数字图像中提取颜色特征依赖于对数字图像中颜色的表示和颜色理论的理解。颜色空间对于相关颜色以数字形式表示是一个很重要的成分，在不同颜色空间之间的

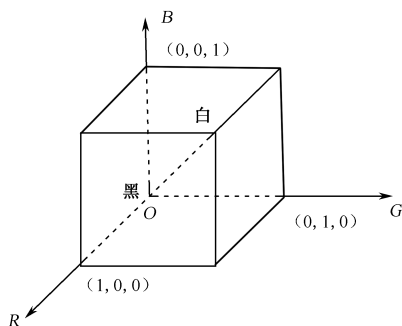


图 2.1 RGB 颜色模型

转换和颜色信息的量化是给定特征提取的决定因素。

### 1. RGB 颜色空间

RGB 颜色空间是图像处理中最基础、最常用的颜色空间，RGB 颜色空间的三维空间包括  $R$ 、 $G$ 、 $B$  三个坐标轴，如图 2.1 所示。我们感兴趣的部分是个立方体，原点对应于黑色；离原点最远的顶点对应于白色；立方体与 3 个坐标轴的交点对应于三基色，即红色、绿色和蓝色；剩余的 3 个顶点对应于三

补色，即品红（即红加蓝）、蓝绿（即绿加蓝）和黄（即红加绿）。在这个模型中，从黑到白的灰度值分布在从原点到离原点最远顶点间的连线上，而立方体内其余各点对应的不同颜色可以用该点到原点的向量来表示。方便起见，可将立方体归一化为单位立方体，因此  $R$ 、 $G$ 、 $B$  的值都在区间  $[0,1]$  上。根据这个模型，每幅彩色图像包括 3 个独立的基色平面，或者说可分解到 3 个平面上。

色觉的产生需要发光光源的光通过反射或透射的方式传递到眼睛，刺激视网膜细胞引起神经信号传输到大脑，然后人脑对此加以解释产生视觉。设组成某颜色  $C$  所需的 3 个刺激量分别用  $X$ 、 $Y$  和  $Z$  表示，3 个刺激量与  $R$ 、 $G$  和  $B$  有如下关系，即

$$\begin{aligned} X &= 0.490R + 0.310G + 0.200B \\ Y &= 0.177R + 0.812G + 0.011B \\ Z &= 0.000R + 0.010G + 0.990B \end{aligned} \quad (2-1)$$

对白光，有  $X=1$ ， $Y=1$ ， $Z=1$ 。设每种刺激量的比例系数为  $x$ 、 $y$ 、 $z$ ，则有  $C = xX + yY + zZ$ 。比例系数  $x$ 、 $y$ 、 $z$  也称为色系数，其定义为

$$\left. \begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z} \end{aligned} \right\} \quad (2-2)$$

可以看出， $x + y + z = 1$ 。

RGB 颜色空间的主要缺点是不直观，从  $R$ 、 $G$ 、 $B$  的值中很难知道该值所表示颜色的认知属性，因此 RGB 颜色空间不符合人对颜色的感知心理。另外，RGB 颜色空间是最不均匀的颜色空间之一，两种颜色之间的知觉差异不能采用该颜色空间中两个颜色点之间的距离来表示。

## 2. HSV 颜色空间

HSV 颜色空间是一种面向视觉感知的颜色模型。从心理学和视觉的角度出发,人眼的色彩知觉主要包括 3 个要素:色调、饱和度和亮度。

色调 ( $H$ ) 是指光的颜色,它与混合光谱中主要光波长相联系,如红、橙、黄、绿、青、蓝、紫分别表示不同的色调。饱和度 ( $S$ ) 是指彩色的深浅程度,即与一定色调的纯度相关。饱和度高表示颜色深,如深红;饱和度低表示颜色浅,如浅红。饱和度的高低与色光中白光成分的多少有关。一种纯彩色光中加入的白光成分越少,其饱和度就越高;反之,白光成分越多,饱和度就越低。因而,饱和度反映了某种色光被冲淡的程度。亮度 ( $V$ ) 是指人眼感受到的光的明暗程度,亮度与物体的反射率成正比,无彩色就是指只有亮度一个维的变化。对彩色来说,颜色中掺入白色越多就越明亮,掺入黑色越多亮度就越小。

用一个三维空间纺锤体可以将色调、饱和度和亮度表示出来,如图 2.2 所示。其中,立体的竖直轴代表黑白系列亮度的变化,圆周上各点代表不同的色调,从圆周向圆心过渡表示饱和度逐渐降低。HSV 颜色空间具有两大特点:其一,亮度分量与图像的彩色信息无关;其二,色调和饱和度分量与人感受颜色的方式是紧密相连的。这些特点使 HSV 颜色空间非常适合于借助人的视觉系统来感知彩色特性的图像处理算法。HSV 颜色空间直接对应于人眼色彩视觉特征的三要素,通道之间各自独立,因此可以独立感知各颜色分量的变化,其中色调尤其影响着人的视觉判断。

因此,在基于内容的图像检索中,应用这种颜色模型会更适合用户的视觉判断。

由于数字图像一般均采用 RGB 颜色模型来显示,因此在进行处理时,需要将颜色值由 RGB 空间转换至 HSV 空间。给定 RGB 颜色空间中的值  $(R, G, B)$ ,  $R, G, B \in [0, 1]$ , 则转换到 HSV 空间的  $H$ 、 $S$ 、 $V$  值计算如下<sup>[2]</sup>。

$$V = \frac{1}{\sqrt{3}}(R + G + B) \quad (2-3)$$

$$S = 1 - \frac{\sqrt{3}}{V} \min(R, G, B) \quad (2-4)$$

$$H = \begin{cases} \theta, & G \geq B \\ 2\pi - \theta, & G < B \end{cases} \quad (2-5)$$

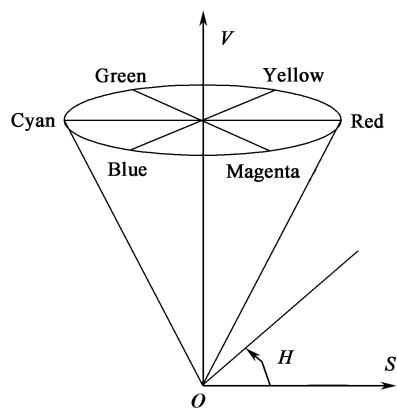


图 2.2 HSV 颜色模型

其中,  $\theta = \arccos \left[ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right]$ 。从 RGB 颜色空间到 HSV 颜色空间的

转换, 还存在一种快速近似的转换公式, 即

$$V = \max(R, G, B) \quad (2-6)$$

$$S = \frac{V - \min(R, G, B)}{V} \quad (2-7)$$

令  $r' = \frac{V-R}{V-\min(R, G, B)}$ ,  $g' = \frac{V-G}{V-\min(R, G, B)}$ ,  $b' = \frac{V-B}{V-\min(R, G, B)}$ , 则有

$$H' = \begin{cases} 5 + b', & R = \max(R, G, B) \text{ 且 } G = \min(R, G, B) \\ 1 - g', & R = \max(R, G, B) \text{ 且 } G \neq \min(R, G, B) \\ 1 + r', & G = \max(R, G, B) \text{ 且 } B = \min(R, G, B) \\ 3 - b', & G = \max(R, G, B) \text{ 且 } B \neq \min(R, G, B) \\ 3 + g', & B = \max(R, G, B) \text{ 且 } R = \min(R, G, B) \\ 5 - r', & \text{其他} \end{cases} \quad (2-8)$$

$$H = 60 \times H' \quad (2-9)$$

由上述公式可知,  $H \in [0^\circ, 360^\circ]$ ,  $S \in [0, 1]$ ,  $V \in [0, 1]$ 。

从 HSV 颜色空间到 RGB 颜色空间的转换公式如下。

(1) 当  $H \in [0^\circ, 120^\circ]$  时,

$$R = \frac{V}{\sqrt{3}} \left[ 1 + \frac{S \cos H}{\cos(60^\circ - H)} \right], \quad B = \frac{V}{\sqrt{3}}(1 - S), \quad G = \sqrt{3}V - R - B \quad (2-10)$$

(2) 当  $H \in [120^\circ, 240^\circ]$  时,

$$G = \frac{V}{\sqrt{3}} \left[ 1 + \frac{S \cos(H - 120^\circ)}{\cos(180^\circ - H)} \right], \quad R = \frac{V}{\sqrt{3}}(1 - S), \quad B = \sqrt{3}V - G - R \quad (2-11)$$

(3) 当  $H \in [240^\circ, 360^\circ]$  时,

$$B = \frac{V}{\sqrt{3}} \left[ 1 + \frac{S \cos(H - 240^\circ)}{\cos(300^\circ - H)} \right], \quad G = \frac{V}{\sqrt{3}}(1 - S), \quad R = \sqrt{3}V - G - B \quad (2-12)$$

### 3. CIEL\*a\*b\*颜色空间和 CIEL\*u\*v\*颜色空间

CIEL\*a\*b\*和 CIEL\*u\*v\*颜色空间是两种均匀的颜色模型。均匀颜色模型本质上仍是面向视觉感知的颜色模型, 只是在视觉感知方面更为均匀。从视觉感知均匀的角度来说, 人所感知到的两种颜色的距离应该与这两种颜色在表达它们的颜色空间中的距离越成比例越好。换句话说, 如果在某一颜色空间中, 人所观察到的两种颜色的区别程度与该颜色空间中两点间的欧氏距离相对应, 则称该空间为均匀的颜色空间<sup>[3]</sup>。

为了将易测的空间距离作为色彩感觉差别量的度量, 1976 年, CIE (国际照明委

员会) 公布了两个标准的同一性空间 CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 颜色空间和 CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色空间。这两个空间是在 CIE 1931 年公布的 XYZ 空间的基础上得到的。RGB 颜色空间到 CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 颜色空间和 CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色空间的转换如下。

RGB→CIEXYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.49 & 0.31 & 0.2 \\ 0.177 & 0.812 & 0.011 \\ 0 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2-13)$$

CIEXYZ→CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup>

$$\begin{aligned} L^* &= 116(Y/Y_n)^{\frac{1}{3}} - 16, \quad Y/Y_n > 0.008856 \\ L^* &= 903.3(Y/Y_n), \quad Y/Y_n \leq 0.008856 \\ a^* &= 500 \left[ (X/X_n)^{\frac{1}{3}} - (Y/Y_n)^{\frac{1}{3}} \right] \\ b^* &= 200 \left[ (Y/Y_n)^{\frac{1}{3}} - (Z/Z_n)^{\frac{1}{3}} \right] \end{aligned} \quad (2-14)$$

CIEXYZ→CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup>

$$\begin{aligned} L^* &= 116(Y/Y_n)^{\frac{1}{3}} - 16, \quad Y/Y_n > 0.008856 \\ L^* &= 903.3(Y/Y_n), \quad Y/Y_n \leq 0.008856 \\ u^* &= 13L^*(u' - u'_n) \\ v^* &= 13L^*(v' - v'_n) \end{aligned} \quad (2-15)$$

其中,  $u' = \frac{4X}{X+15Y+3Z}$ ,  $v' = \frac{9Y}{X+15Y+3Z}$ ;  $(X_n, Y_n, Z_n)$  表示参考白色对应的值;  
 $u'_n$  和  $v'_n$  表示参考白色的变换值, 其定义与  $u'$  和  $v'$  的定义相同。

CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 和 CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色空间均是基于对立色理论。CIEL<sup>\*</sup>u<sup>\*</sup>v<sup>\*</sup> 颜色模型与设备无关, 适用于显示器显示和根据加色原理进行组合的场合, 该模型比较强调对红色的表示, 即对红色的变化比较敏感, 但对蓝色表示比较粗糙。CIEL<sup>\*</sup>a<sup>\*</sup>b<sup>\*</sup> 颜色模型也与设备无关, 适用于接近自然光照明的场合, 该模型比较强调对绿色的表示 (对绿色比较敏感), 其次是红色和蓝色。

#### 4. YCrCb 颜色空间

YCrCb 颜色空间是一种用于数字图像的颜色标准。该模型中 Y 代表了光源的亮度, 色度信息组合在 Cr、Cb 中, 其中, Cr 代表了光源中的红色分量, Cb 代表了光源中的蓝色分量。

亮度给出了颜色亮或暗的程度信息, 是人在观察光照时感知亮度变化的心理上的度量, 可通过特定照明中的强度成分的加权和来计算。在 RGB 光源中, 光源的绿色

分量对亮度影响最大，蓝色分量对亮度影响最小，因此亮度公式一般可表示为

$$Y=0.299 R +0.587G +0.114B \quad (2-16)$$

由于人眼对于亮度的敏感程度大于对于色度的敏感程度，所以完全可以让相邻的像素使用同一个色度值，而人眼的感觉不会起太大的变化，从而通过损失色度信息来达到节省存储空间的目的。因此，YCrCb 颜色模型适合于图像文件的压缩，目前它被多种图像文件格式所采用，如 JPEG、MPEG 等国际标准。

RGB 颜色空间与 YCrCb 颜色空间的转换关系为

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.144 \\ 0.5 & -0.4187 & -0.0813 \\ -0.1687 & -0.3313 & 0.5 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (2-17)$$

## 2.1.2 颜色量化

颜色量化是图像工程中的一项基本而重要的技术，自然界具有无限丰富的颜色，真彩色图像具有  $2^{24}$  种颜色，而在某些情况下（如图像检索、印染等）对颜色的数目有一定的限制。如何选取有代表性的若干种颜色，并把各种颜色都归并到这些代表色上，就是颜色量化需要解决的问题。

### 1. 颜色量化的定义

颜色量化可形式化表述如下<sup>[4]</sup>：设  $\mathbf{c}_i$  是颜色空间中的一个三维向量， $\mathbf{C} = \{\mathbf{c}_i | i=1,2,\dots,N\}$  表示输入图像中颜色的集合（ $N$  表示颜色的数目）， $\overline{\mathbf{C}} = \{\overline{\mathbf{c}}_j | j=1,2,\dots,K\}$ （ $K \leq N$ ）表示输出图像中颜色的集合，颜色量化是一个映射过程，即

$$q: \mathbf{C} \rightarrow \overline{\mathbf{C}} \quad (2-18)$$

颜色量化遵循距离最近准则：输入图像中的任一颜色  $\mathbf{c}$  将被映射到调色板  $\overline{\mathbf{C}}$  中距离最近的颜色  $\overline{\mathbf{c}}$ ，即

$$\overline{\mathbf{c}} = q(\mathbf{c}) = \|\mathbf{c} - \overline{\mathbf{c}}\| = \min_{j=1,2,\dots,K} \|\mathbf{c} - \overline{\mathbf{c}}_j\| \quad (2-19)$$

同时，在颜色集合  $\mathbf{C}$  中得到  $K$  个聚类  $S_k$ （ $k=1,2,\dots,K$ ），即

$$S_k = \{\mathbf{c} \in \mathbf{C} | q(\mathbf{c}) = \overline{\mathbf{c}}_k\} \quad (2-20)$$

其中， $\overline{\mathbf{c}}_k$  为  $K$  个聚类的聚类中心，它们组成输出图像的调色板。

由于颜色量化是把非常丰富的颜色量化到较少的颜色上去，因此不可避免地存在偏差，该过程是一个有损的过程。颜色量化能否取得理想的效果，关键在于能否解决输入图像的整体层次和局部细节之间的矛盾，一个好的颜色量化算法需要在这对矛盾中找到合适的平衡点。

## 2. 常用的颜色量化方法

目前常用的颜色量化方法大体可以分为分割算法和聚类算法两大类<sup>[5]</sup>。分割算法的基本思想是将图像中出现的频率最高的  $K$  种色彩作为调色板, 然后将其余颜色按照距离最近准则映射到调色板中, 此类方法重构图像的层次感较丰富, 但会丢失出现频率低的色彩, 因而无法保留细节, 使局部模糊。代表性的分割算法有频度序列法<sup>[6]</sup>、八叉树法<sup>[7]</sup>等。聚类算法则先选择若干聚类中心, 然后按某种准则对颜色进行迭代聚合, 直到合适的分类为止。典型的有  $k$  均值聚类算法<sup>[8]</sup>、模糊  $c$  均值聚类算法<sup>[9]</sup>等。聚类算法为近似最优算法, 但需迭代运算, 计算量大, 而且量化结果往往依赖于初始聚类中心的选取。另外, 聚类算法容易将相近的色彩合并, 而破坏色彩的层次感。

除了上述的量化方法外, 还存在其他许多颜色量化方法。如 QBIC 系统将 RGB 空间初始量化为 163 个单元, 每个单元中心转化为 Munsell 颜色空间, 用户查询时可根据需要选择直方图维数, 默认是 64 维。VisualSEEK 系统将 HSV 空间划分为 166 份, 其中, hue 划分为 18 份, value 和 saturation 各 3 份, 再加上灰度空间 4 份, 这种量化方法与 MPEG-7 中尺度描述子 (scalable color descriptor) 标准相同。

Androutsos 等人<sup>[10]</sup>通过实验对 HSV 颜色空间进行了大致的划分, 即将亮度大于 75% 且饱和度大于 20% 的区域定义为亮彩色区域, 将亮度大于 75% 且饱和度小于 20% 的区域定义为白色区域, 将亮度小于 20% 的区域定义为黑色区域, 这样在每个范围内再进行细分, 从而将 HSV 颜色空间划分为 29 种颜色, 这种方法从理论上与人的感知更加一致, 对该方法的改进见文献[11]、[12]。其他的颜色量化方法有改进的  $k$  均值的量化方法<sup>[13]</sup>、基于分层聚类的方法<sup>[14]</sup>、基于自组织网络的量化技术<sup>[15]</sup>、基于 GRM (Generic Roughness Measure) 的量化方法<sup>[16]</sup>、基于软计算技术的量化方法<sup>[17]</sup>等。

### 2.1.3 全局颜色特征

总体来看, 颜色特征的表达主要集中在两个方面: 全局颜色特征和空间颜色特征。本小节将主要讨论全局颜色特征的描述方法。

#### 1. 颜色直方图

给定一幅图像  $(f_{xy})_{M \times N}$ ,  $f_{xy}$  表示像素点  $(x, y)$  处的颜色值,  $M \times N$  表示图像的尺寸, 图像所包含的颜色集记为  $C$ , 则图像的颜色直方图可表示为

$$h_c = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f_{ij} - c), \quad \forall c \in C \quad (2-21)$$

颜色直方图作为基于颜色特征进行图像检索的一种重要的方法, 具有特征提取和相似度计算简便, 并且随图像尺度、旋转等变化不敏感的特点。

但颜色直方图也存在许多缺点。

(1) 颜色直方图描述的是图像颜色的统计特性, 丢失了颜色的空间分布信息, 因此, 对两个颜色直方图相似的图像来说, 如果颜色的空间分布差别很大, 图像的内容会很不相同。

(2) 存在特征维数过高的问题。

(3) 由于对图像颜色的量化处理, 可能将视觉不同的颜色量化到同一区间, 也可能将视觉相同的颜色量化到不同的区间, 因而造成误检现象。

针对上述问题, 近年来许多改进的算法被提了出来, 如累加直方图方法和局部累加直方图<sup>[18, 19]</sup>、模糊直方图<sup>[20-22]</sup>、颜色差异直方图<sup>[23]</sup>、主色直方图<sup>[24-26]</sup>、主色调直方图<sup>[27]</sup>、颜色矢量角直方图<sup>[28]</sup>等。

## 2. 颜色不变量

Funt 及 Finlayson<sup>[29]</sup>认为光照度在一定的区域内可以看作常量。他们对颜色值的对数求导 (拉普拉斯或方向导数), 计算直方图, 得到的实际是相邻颜色边界的长度, 以颜色变化率为索引, 以此排除光照成分对颜色的影响。这种方法当图像库在均匀光照下比颜色直方图的方法略差一些, 但当图像库中的图像受到光照影响时要好一些。Gevers 等人<sup>[30]</sup>从 Shafer 双色反射模型, 推导出了受光照影响小的颜色模型, 如式(2-22)所示。Gijsenij 等<sup>[31]</sup>对近年来颜色不变量的提取方法进行了分析和评价。

$$\left. \begin{aligned} l_1 &= \frac{(R-G)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_2 &= \frac{(R-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \\ l_3 &= \frac{(G-B)^2}{(R-G)^2 + (R-B)^2 + (G-B)^2} \end{aligned} \right\} \quad (2-22)$$

## 3. 颜色矩

Stricker 和 Orengo 所提出的颜色矩 (color moments)<sup>[18]</sup>是另一种非常简单且有效的颜色特征描述方法, 这种方法的数学基础在于图像中任何颜色分布均可以用它的矩来表示。此外, 由于颜色分布信息主要集中在低阶矩中, 因此仅采用颜色直方图特征的一阶矩 (mean)、二阶中心矩 (variance) 和三阶中心矩 (skewness) 就足以表达图像的颜色特征。与颜色直方图相比, 该方法的另一个好处在于无须对颜色进行量化且降低了颜色特征的维数。颜色 3 个低阶矩的数学表达形式为

$$\mu_i = \frac{1}{n} \sum_{j=1}^n h_{ij} \quad (2-23)$$



$$\sigma_i = \left[ \frac{1}{n} \sum_{j=1}^n (h_{ij} - \mu_i)^2 \right]^{\frac{1}{2}} \quad (2-24)$$

$$s_i = \left[ \frac{1}{n} \sum_{j=1}^n (h_{ij} - \mu_i)^3 \right]^{\frac{1}{3}} \quad (2-25)$$

其中,  $h_{ij}$  表示第  $i$  个颜色通道分量中灰度为  $j$  的像素出现的概率;  $n$  表示灰度级数。因此, 图像的颜色矩一共有 9 个分量 (3 个颜色分量, 每个颜色分量有 3 个低阶矩)。颜色矩同其他颜色特征相比是非常简洁的, 但实验发现图像低阶矩的检索效率比颜色直方图的检索效率要低。在实际应用中为了避免低阶矩较弱的分辨能力, 往往将颜色矩同其他图像特征联合应用, 在利用其他特征进行图像检索前, 可首先采用颜色矩过滤, 缩小检索范围。

#### 4. 颜色熵

根据颜色直方图特性和信息论中信息熵的概念, John<sup>[32]</sup>提出采用图像颜色的信息熵 (简称颜色熵) 来表示图像的颜色特征, 从而将图像的颜色直方图由多维降低到一维。设图像的归一化颜色直方图表示为  $(h_1, h_2, \dots, h_n)$ , 如果我们将图像的颜色直方图看作图像中不同颜色的像素在图像空间中出现的概率密度函数, 则根据信息熵理论, 图像颜色的信息熵可表示为

$$E = - \sum_{i=1}^n h_i \log_2(h_i) \quad (2-26)$$

虽然采用熵的方法有效地降低了图像直方图特征的维数, 但在利用颜色熵进行图像检索时, 其分辨能力是较低的。因此, 颜色熵特征往往也需要和其他图像特征相结合进行检索, 在利用其他图像特征进行图像检索前, 首先利用颜色熵缩小检索范围。

#### 5. 改进的颜色熵及颜色矩

在采用颜色熵表示图像的颜色特征时, John 并没有考虑熵的对称性对这一特殊应用的影响, 即向量各分量的次序任意改变时, 熵值不变, 熵函数的取值只与向量的概率分布有关。同时, 对颜色矩来说, 也存在类似的问题, 颜色矩表现出的也是一种统计特征, 视觉不同的颜色直方图如果具有相似的概率分布, 也将具有相似的颜色矩。

如图 2.3 所示的 3 个直方图  $H_1$ 、 $H_2$  及  $H_3$ , 显然  $H_1$  和  $H_2$  具有一定的视觉相似性, 它们与  $H_3$  的视觉特性完全不同。但由于这 3 个直方图具有相同的概率分布特征, 因此它们具有相同的颜色熵和矩。为了消除这种对称性对图像检索结果造成的影响及考虑到人的视觉特性, 文献[27]提出了直方图排序法、直方图面积法、直方图排序法和直方图面积法的线性组合来消除这种影响。

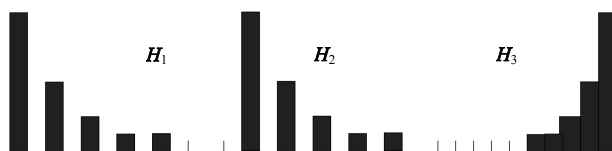


图 2.3 颜色直方图示例

### 1) 直方图排序法

虽然图 2.3 所示的 3 个直方图具有相同的概率分布特征，但很明显，不同概率在直方图中的位置完全不同，我们可借助这种位置特征来进行区分。

设  $H$  表示某幅图像的颜色直方图，我们采用某种排序方法（如冒泡排序）将直方图  $H$  的各个分量从大到小（或从小到大）进行排序，假设排序后的颜色直方图为  $H'$ ，在排序过程中直方图各分量移动的总次数记为  $M_H$ 。由颜色熵和矩的特性可知，直方图  $H$  和  $H'$  具有相同的颜色熵和矩，因此这种排序不会造成直方图颜色熵和矩的改变。对具有相同或相近概率分布的直方图来说，如果直方图中各分量的次序不同，在对直方图进行排序后，各分量的移动次数也将不同，因此采用排序时各分量移动次数的差异可有效地消除这种影响。

按照以上分析，我们在计算颜色熵和矩时引入了加权函数  $f_1(H)$ ，即

$$f_1(H) = 1 + M_H / M_{\max} \quad (2-27)$$

其中， $M_{\max}$  表示排序时直方图分量需要移动的最大次数，即对逆序直方图进行排序时直方图各分量移动的次数。

从而改进的图像颜色熵和矩可分别表示为

$$E(H) = -f_1(H) \sum_{i=1}^n h_i \log_2(h_i) \quad (2-28)$$

$$\left. \begin{aligned} \mu &= f_1(H) \frac{1}{N} \sum_{i=1}^N p_i \\ \sigma &= f_1(H) \left[ \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^2 \right]^{\frac{1}{2}} \\ s &= f_1(H) \left[ \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^3 \right]^{\frac{1}{3}} \end{aligned} \right\} \quad (2-29)$$

设  $E_0$  代表图 2.3 所示直方图的颜色熵， $\mu_0$ 、 $\sigma_0$  及  $s_0$  表示它们的颜色矩。如果我们按照从小到大的顺序将 3 个直方图从小到大进行排列，则  $M_{H_1} = 25$ ， $M_{H_2} = 20$ ， $M_{H_3} = 0$ ， $M_{\max} = 45$ 。如果我们采用  $L_1$  距离进行度量，对直方图颜色熵来说有

$$d_{L_1}(H_1, H_3) = 0.56E_0 > d_{L_1}(H_2, H_3) = 0.44E_0 > d_{L_1}(H_1, H_2) = 0.11E_0 \quad (2-30)$$

对颜色矩来说有

$$\begin{aligned}
d_{L_1}(\mathbf{H}_1, \mathbf{H}_3) &= 0.56(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\
&> d_{L_1}(\mathbf{H}_2, \mathbf{H}_3) = 0.44(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\
&> d_{L_1}(\mathbf{H}_1, \mathbf{H}_2) = 0.11(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0)
\end{aligned} \tag{2-31}$$

其中,  $\omega_i$  ( $1 \leq i \leq 3$ ) 代表权值。

显然, 从上述式子的计算结果可以看出, 直方图  $\mathbf{H}_1$  和  $\mathbf{H}_2$  间的距离较小, 而它们与直方图  $\mathbf{H}_3$  的距离较大, 该结果也与人类的视觉特性一致。

## 2) 直方图面积法

在采用直方图排序法进行改进时, 仅仅考虑了直方图中各分量在整个向量中的位置因素, 并没有考虑该位置的分量所对应概率的大小, 为此, 我们进一步提出了直方图面积法。定义直方图的面积为

$$A_{\mathbf{H}} = \sum_{i=1}^n (h_i \times i) \tag{2-32}$$

由上式可知, 当  $p_n = 1$  时, 直方图的面积达到最大值  $A_{\max} = n$ 。对具有相同或相近概率分布的直方图来说, 如果直方图中各分量的次序不同, 则直方图所对应的面积一般也是不同的, 因此采用直方图面积法也可有效地消除熵的对称性的影响。为此, 在计算直方图的颜色熵时, 可引入加权函数  $f_2(\mathbf{H})$ ,  $f_2(\mathbf{H}) = 1 + \frac{A_{\mathbf{H}}}{A_{\max}}$ , 从而图像颜色熵和矩可分别表示为

$$E(\mathbf{H}) = -f_2(\mathbf{H}) \sum_{i=1}^n h_i \log_2(h_i) \tag{2-33}$$

$$\left. \begin{aligned}
\mu &= f_2(\mathbf{H}) \frac{1}{N} \sum_{i=1}^N p_i \\
\sigma &= f_2(\mathbf{H}) \left[ \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^2 \right]^{\frac{1}{2}} \\
s &= f_2(\mathbf{H}) \left[ \frac{1}{N} \sum_{i=1}^N (p_i - \mu)^3 \right]^{\frac{1}{3}}
\end{aligned} \right\} \tag{2-34}$$

设图 2.3 中,  $\mathbf{H}_1 = (1/2, 0, 1/4, 0, 1/8, 0, 1/16, 0, 1/16, 0)$ ,

$\mathbf{H}_2 = (0, 1/2, 0, 1/4, 0, 1/8, 0, 1/16, 0, 1/16)$ ,

$\mathbf{H}_3 = (0, 0, 0, 0, 1/16, 1/16, 1/8, 1/4, 1/2)$ 。

则有  $A_{\max} = 10$ ,  $A_{\mathbf{H}_1} = 2.88$ ,  $A_{\mathbf{H}_2} = 3.88$ ,  $A_{\mathbf{H}_3} = 9.06$ 。采用直方图面积法, 对颜色熵来说有  $d_{L_1}(\mathbf{H}_1, \mathbf{H}_2) = 0.1E_0$ ,  $d_{L_1}(\mathbf{H}_1, \mathbf{H}_3) = 0.62E_0$ ,  $d_{L_1}(\mathbf{H}_2, \mathbf{H}_3) = 0.52E_0$ 。对颜色矩来说有

$$\begin{aligned}
 d_{L_1}(H_1, H_3) &= 0.62(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\
 &> d_{L_1}(H_2, H_3) = 0.52(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0) \\
 &> d_{L_1}(H_1, H_2) = 0.10(\omega_1\mu_0 + \omega_2\sigma_0 + \omega_3s_0)
 \end{aligned}
 \tag{2-35}$$

从计算结果可以看出, 该结果也同人的视觉感知一致。同时, 采用直方图面积法, 直方图  $H_1$  及  $H_2$  间的距离变得更小了, 而它们与直方图  $H_3$  的距离增大了, 这一结果与人类的视觉特性更为接近, 因此利用直方图面积法比直方图排序法取得更好的效果。这是由于直方图面积法既考虑了直方图中各分量的位置因素, 又考虑了该位置的分量所对应概率的大小。

### 3) 直方图排序法和直方图面积法的线性组合

虽然实验中发现采用直方图面积法的改进效果总体上优于直方图排序法, 但对于一些特殊的情况, 直方图面积法的改进效果反而比不上直方图排序法。例如, 对图 2.4 所示的两个直方图来说, 它们具有相同的信息熵, 并且存在  $p_1m_1 + p_2m_2 = p_2m_3 + p_1m_4$ ,

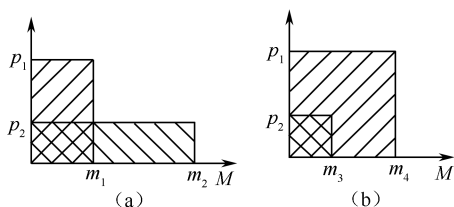


图 2.4 特殊直方图

即具有相同的直方图面积, 因此采用直方图面积法是不能区分出直方图 (a) 和 (b) 的, 但是如果采用直方图排序法, 则可有效地区分出直方图 (a) 和 (b)。

为此, 我们可以进一步采用直方图排序法和直方图面积法的线性组合来进一步增强图像的检索效果。采用两种方法线性组合的权函数可表示为

$$f_3(H) = \gamma_1 f_1(H) + \gamma_2 f_2(H) \tag{2-36}$$

其中,  $\gamma_1$ 、 $\gamma_2$  为权值系数, 且  $\gamma_1 \in [0,1]$ ,  $\gamma_2 \in [0,1]$ ,  $\gamma_1 + \gamma_2 = 1$ 。

由于直方图排序法仅仅考虑了直方图中各分量的位置因素, 并没考虑该位置的分量所对应概率的大小, 而直方图面积法既考虑了直方图中各分量的位置因素, 又考虑了该位置的分量所对应概率的大小, 从而直方图面积法在一定程度上优于直方图排序法。因此, 对于式 (2-36) 在选择权值时, 往往取  $\gamma_1 > \gamma_2$ 。上述 3 种改进方法的实验对比效果见文献[27]。

## 2.1.4 空间颜色特征

上述颜色直方图、颜色矩和颜色熵等所描述的颜色特征是图像的全局颜色特征, 不包括图像颜色的空间分布特征, 因此仅仅利用这些特征进行图像检索时极易造成误检现象。本节主要介绍常用的空间颜色特征的描述方法。

### 1. 颜色直方图的改进

为了在图像的颜色直方图中包含颜色的空间分布信息,文献[33]将边缘信息融入图像的颜色直方图中<sup>[33]</sup>。Guoping Qiu 等人<sup>[34]</sup>根据人的视觉特性与频域中频率的关系,提出采用滤波器将图像的亮度分量转化为不同的频率分层,并进一步将彩色图像转化为不同的频率分层,然后对于每一频率分层采用颜色直方图方法来进行索引,从而有效地提高图像的检索效果。宋擒豹等<sup>[35]</sup>提出了颜色-位置直方图方法,该直方图在不失传统直方图稳健性的前提下,将图像的颜色信息和空间位置信息有机地融合起来,该颜色直方图在反映颜色统计信息的同时也记录了颜色分段虚拟边界的位置信息,因而较好地解决了传统直方图存在的问题。

### 2. 颜色聚合向量

针对颜色直方图和颜色矩等无法描述图像颜色的空间分布信息,Pass 等人<sup>[36]</sup>提出了颜色聚合向量(color coherence vector)的方法。其核心思想是将属于颜色直方图每一个区间内的像素分为两部分,如果该区间内的某些像素所占据的连续区域的面积大于给定的阈值,则将该区域内的像素作为聚合像素,否则作为非聚合像素。假设 $\alpha_i$ 和 $\beta_i$ 分别代表直方图的第 $i$ 个区间中聚合像素和非聚合像素的数目,图像的颜色聚合向量可表示为

$$\langle (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_n, \beta_n) \rangle \quad (2-37)$$

而 $\langle \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n \rangle$ 就是该图像的颜色直方图。由于颜色聚合向量包含了图像颜色的空间分布信息,因而取得了相对于颜色直方图更好的检索效果。

### 3. 颜色相关图

颜色相关图<sup>[37]</sup>(color correlogram)是利用图像中像素间的颜色关系来描述图像颜色空间分布的另一种表达方式。颜色相关图不但刻画了某一种颜色的像素占整个图像的比例,还反映了不同颜色对之间的空间相关性。对于任意的图像 $I$ ,假设 $I_{c(i)}$ 表示图像中颜色为 $c(i)$ 的所有像素的集合,则颜色相关图可以表示为

$$r_{i,j}^{(k)} = \Pr_{p_1 \in I_{c(i)}, p_2 \in I_{c(j)}} [p_2 \in I_{c(j)}, |p_1 - p_2| = k \mid p_1 \in I_{c(i)}] \quad (2-38)$$

其中, $i, j \in \{1, 2, \dots, n\}$ ,  $n$ 表示图像颜色级数; $k \in \{1, 2, \dots, d\}$ ,  $d$ 表示在计算图像的颜色相关图时所设定的像素间的最大距离; $|p_1 - p_2|$ 表示像素 $p_1$ 和 $p_2$ 间的距离。此时,颜色相关图可以看作一张用颜色对 $(i, j)$ 索引的表,其中 $(i, j)$ 的第 $k$ 个分量表示图像中颜色为 $c(i)$ 和 $c(j)$ 且距离为 $k$ 的像素对出现的概率。如果考虑到任何颜色间的相关性,颜色相关图就会变得非常复杂和庞大[空间复杂度为 $O(n^2 d)$ ]。一种简化的方法是颜色自相关图(color auto-correlogram),它仅仅考虑图像空间中具有相同颜色的像素间的空间关系,因此空间复杂度降为 $O(nd)$ 。

#### 4. 局部颜色特征

目前,从划分局部区域的角度来说,常用的划分方法包括基于固定块的图像分割、基于手工的区域分割、采用交互半自动的区域分割及一些自动的颜色分割方法。Hsu等<sup>[38]</sup>提出从图像中选择一些代表颜色,然后将图像划分为矩形区域,每个区域以一种主要的单一颜色作为代表,两幅图像之间的相似性采用两幅图像之间具有相似颜色区域的重叠程度来表示。文献[39]利用固定尺寸的网格将图像划分为不同的区间,并提取每一区间的颜色直方图作为图像的索引,但采用该方法,即使图像中不包含某些颜色,局部颜色直方图特征中仍包含该颜色特征,因而存储代价十分昂贵。为了改变这种情况,Stehling<sup>[40]</sup>提出了改进的方法CSH(Color-Shape Histograms),它统计不同颜色在每一个局部分块的统计特征,这样如果图像中不包含某种颜色,就不必要进行存储。

但是,一幅图像往往具有若干表示主题内容的主题画面,由于对原图像事先缺乏任何先验知识,因此无法确定这些主题画面在图像中的位置和大小。在进行空间划分时尽可能将表现图像内容的主题画面划分到同一分块内是一种好的选择,采用固定网格的划分方法是一种刚性的划分,往往会将图像主题信息划分到不同的网格中。基于此问题,何清法等<sup>[41]</sup>提出采用多分辨率的划分方法,采用不同分辨率的网格将图像划分为不同等级,并且在进行图像划分时,各分块间可采用重叠的方式。文献[42]、[43]等提出空间颜色直方图来描述图像颜色特征。文献[44]、[45]等提出基于感兴趣点颜色特征的图像检索方法,从而有效包含图像的空间颜色特征。

#### 5. 环形颜色直方图与空间分布熵

在文献[46]、[47]中,提出利用信息熵来描述图像颜色的空间特征,并提出了相应的增强算法,具体介绍如下。

##### 1) 环形颜色直方图

Rao等<sup>[42]</sup>提出采用环形颜色直方图(annular color histogram)来描述图像颜色的空间分布特征。设 $I$ 表示任意一幅图像, $I(x,y)$ 表示像素 $(x,y)$ 处的颜色值,则 $A_i = \{(x,y) | (x,y) \in I, I(x,y) = i, 1 \leq i \leq n\}$ 表示图像中颜色为 $i$ 的所有像素的集合,其中 $n$ 表示图像颜色的量化级数。设 $|A_i|$ 表示集合 $A_i$ 中像素的数目, $O_i = (x_i, y_i)$ 为图像中颜色为 $i$ 的所有像素的质心, $x_i$ 和 $y_i$ 定义为

$$x_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} x, \quad y_i = \frac{1}{|A_i|} \sum_{(x,y) \in A_i} y \quad (2-39)$$

设 $r_i$ 表示图像中颜色为 $i$ 的像素同其质心的最大距离,其定义为

$$r_i = \max_{(x,y) \in A_i} (\sqrt{(x-x_i)^2 + (y-y_i)^2}) \quad (2-40)$$

对于给定的一个正整数 $N$ ,把 $r_i$ 分为 $N$ 等份,然后以质心 $O_i$ 为圆心,以 $(j \times r_i)/N$

为半径 ( $1 \leq j \leq N$ ) 画圆可得到  $N$  个环。那么由内至外每个环和  $A_i$  的交点把  $A_i$  分为  $A_{i1}, A_{i2}, \dots, A_{iN}$ , 我们称  $A_{i1}, A_{i2}, \dots, A_{iN}$  为  $A_i$  的一个分割。设  $|A_{ij}|$  表示环形区间  $j$  内颜色为  $i$  的像素数目, 则  $(|A_{i1}|, |A_{i2}|, \dots, |A_{iN}|)$  就构成了颜色  $i$  的环形颜色直方图。由于质心  $O_i$  具有平移和旋转不变性, 所以求取的环形颜色直方图同样具有平移和旋转不变性。

但是, 该直方图与图像中某一环形区间内相同颜色像素的数目相关, 因此该环形颜色直方图受图像的尺寸影响较大, 不具有尺度不变特性。为了使环形颜色直方图满足尺度不变性, 我们采用如下的方法对环形颜色直方图进行了归一化处理

$$p_{ij} = \frac{|A_{ij}|}{|A_i|} \quad (2-41)$$

这样, 经过归一化后的环形颜色直方图满足旋转、平移和尺度不变性。

## 2) 空间分布熵

通过上述处理, 图像中的每一种颜色均对应一个环形颜色直方图, 若直接采用环形颜色直方图结合图像的颜色直方图来进行图像检索, 由于每一种颜色均对应一个环形颜色直方图, 因此这将大大增加存储这些特征所需的存储空间, 同时还会造成检索速度的下降。为此, 我们利用熵的特性, 提出采用颜色空间分布熵来描述颜色的空间分布特征。颜色  $i$  的空间分布熵表示为

$$e_i = -\sum_{j=1}^N p_{ij} \log_2(p_{ij}) \quad (2-42)$$

空间分布熵反映了具有某种颜色的像素在图像空间中的平均分散程度, 颜色空间分布熵越大, 表明具有该颜色的像素在图像空间中的分布越分散, 否则, 表明具有该颜色的像素在图像空间中的分布越集中。因此, 采用颜色空间分布熵可有效地表征颜色的空间分布特征, 颜色特征的维数也将大大降低。同时, 由于归一化后的环形颜色直方图满足平移、旋转和尺度不变性, 因此颜色空间分布熵也满足平移、旋转和尺度不变性。

## 3) 加权空间分布熵

在人眼看来, 对具有相同颜色的像素在空间的分布来说, 若该颜色的像素在空间中的分布越集中, 则该种颜色对人眼的视觉刺激越大; 相反, 若该颜色的像素在空间中的分布越分散, 则对人眼的视觉刺激越小。对空间分布熵来说, 熵越大, 表明具有某颜色的像素在图像空间中的分布越分散, 因此对人眼的视觉刺激越小; 熵越小, 表明具有某颜色的像素在图像空间中的分布越集中, 因此对人眼的视觉刺激越大。

一般, 距离质心越近的环形区间, 其像素分布越集中; 距离质心越远的环形区间, 其像素分布相对分散。为此, 在计算颜色空间分布熵时, 引进权函数  $f_4(j)$  ( $j$  表示不同的环形区间) 来反映不同环形区间对图像内容 (人眼视觉的刺激程度) 的贡献程度。

结合人类的视觉特征及熵的特性，权函数的设置满足如下规则：距离质心较近的区间应赋予较小的权值，距离质心较远的区间应赋予较大的权值。从而权函数定义为

$$f_4(j) = 1 + \frac{j}{N} \quad (2-43)$$

图像的加权空间分布熵表示为

$$e_i = - \sum_{j=1}^N f_4(j) p_{ij} \log_2(p_{ij}) \quad (2-44)$$

由于不同的环形区间引入不同的权值，这样就消除了由于空间分布熵相近而空间分布直方图不同对检索结果带来的影响，该方法也可以有效地解决熵的对称性对图像检索结果所带来的影响。

## 6. 位平面与位平面熵

由前面的分析可知，颜色熵反映了图像的全局统计信息，丢弃了图像的空间分布信息，具有相同信息熵的图像可能在视觉上是完全不同的，所以图像的全局信息熵不足以反映出图像间的差异。为此，这里引入位平面熵的概念<sup>[28]</sup>，使得图像特征能用一个熵矢量来描述，同时，对于熵相同的两幅图像所提取的位平面熵却不同，从而解决熵相同而图像内容不同的问题。

### 1) 位平面

将一个像素的灰度值分解为二进制值，所有同权值的位（0 或 1）构成的平面叫位平面。例如，一幅灰度为 256 的图像，每个像素占一个字节，即 8 个二进制位，按从高位到低位的排列为  $b_7b_6b_5b_4b_3b_2b_1b_0$ ，那么所有像素的  $b_0$  位就构成第 0 个位平面， $b_1$  位就构成第 1 个位平面，以此类推，将图像分解为 8 个位平面，这样一幅灰度图像就可以被看作 8 个位平面的叠加，每个位平面被看作一幅二值图像。

设  $I$  表示一幅灰度图像， $I(i, j)$  为其中的一个像素，则对该像素的位平面分解定义为

$$g_t(i, j) = B_t[I(i, j)] = \begin{cases} 1, & \left[ \text{Int}\left(\frac{I(i, j)}{2^t}\right) \right] \text{MOD } 2 = 1 \\ 0, & \left[ \text{Int}\left(\frac{I(i, j)}{2^t}\right) \right] \text{MOD } 2 = 0 \end{cases} \quad (2-45)$$

其中， $0 \leq t \leq 7$ ， $B_t(\bullet)$  代表图像的位平面分解操作，显然  $g_t(i, j) \in \{0, 1\}$ 。经过这样的处理后，该图像就被分解为  $g_0, g_1, \dots, g_7$ ，即完成了对图像进行位平面分解的操作。反过来，图像像素  $I(i, j)$  的合成可以表示为

$$I(i, j) = \sum_{t=0}^7 g_t(i, j) \times 2^t \quad (2-46)$$

一幅 8 位灰度图像的 8 个位平面分解结果如图 2.5 所示，其中，图 2.5 (a) 是一



幅示例图像，图 2.5 (b) 到图 2.5 (i) 是它的 8 个位平面（从高位面到低位面）。从图中可以看出，每个位平面都能够反映图像的频率和方向在局部范围内的变化强度，较高位平面包含了大多数在视觉上很重要的数据，其他位平面则包含了图像中更多的细节信息。低位平面图比高位平面图复杂，包括的细节多，也更随机。显然，对位平面而言，越高的位平面包含的信息就越重要，越低的位平面由于其杂乱无章而表现出一定的自相似性。

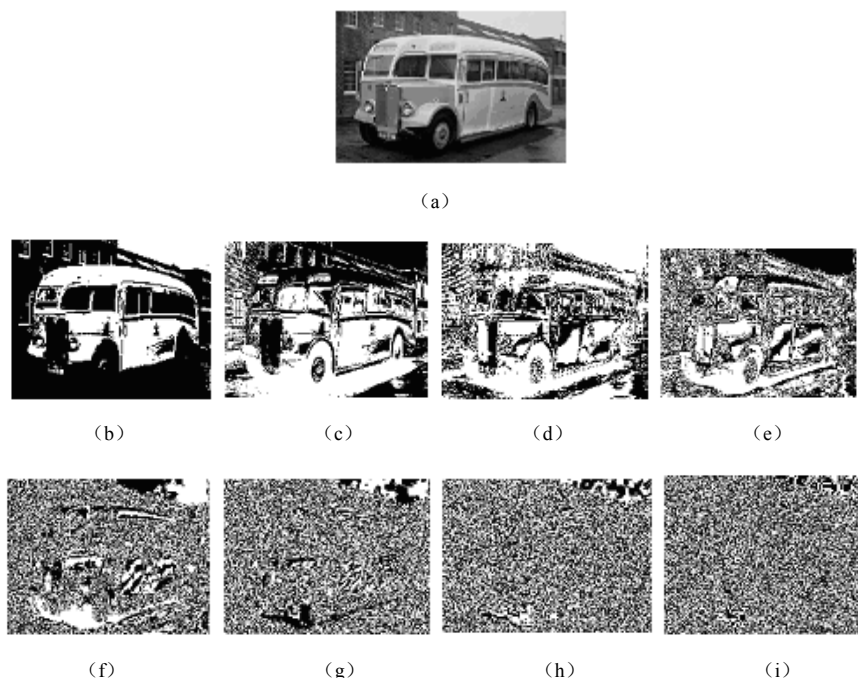


图 2.5 示例图像及其 8 个位平面

采用这种位平面的表示方法存在一个缺点，即像素点灰度值的微小变化会对位平面的复杂度产生较明显的影响。例如，当空间相邻的两个像素的灰度值分别为 127 ( $01111111_2$ ) 和 128 ( $10000000_2$ ) 时，图像的每个位平面上在这个位置处都会有从 0 到 1（或从 1 到 0）的过渡。为减小这种影响，算法中采用灰度码的方法来表示位平面。图像的灰度码可由下式计算。

$$G_i = \begin{cases} g_i \oplus g_{i+1}, & 0 \leq i \leq m-2 \\ g_i, & i = m-1 \end{cases} \quad (2-47)$$

其中， $\oplus$  代表异或操作； $g_i$  表示位平面分解得到的第  $i$  个位平面； $G_i$  指位面  $g_i$  的灰度码表示。这种码的独特性质是相连的码字只有 1 个比特位的区别，这样，像素点灰度值的小变化就不会影响所有的位平面。而且对于每个用其相应的灰度码来表示的二进制位平面，其灰度码是唯一的，反之亦然。仍考虑上述空间相邻的两个像素的灰度值

分别为 127 和 128 的例子，若用式 (2-47) 的灰度码来表示的话，则这里只有位平面 7 有从 0 到 1 的一个过渡，其他位平面没有变化，此时对应 127 和 128 的灰度码分别是  $01000000_2$  和  $11000000_2$ 。图 2.6 给出了图 2.5 (a) 用灰度码表示的位平面图 (从高位到低位)。可以看出，用灰度码表示的位平面图复杂度较低，但具有视觉意义信息的位平面图数量更多。

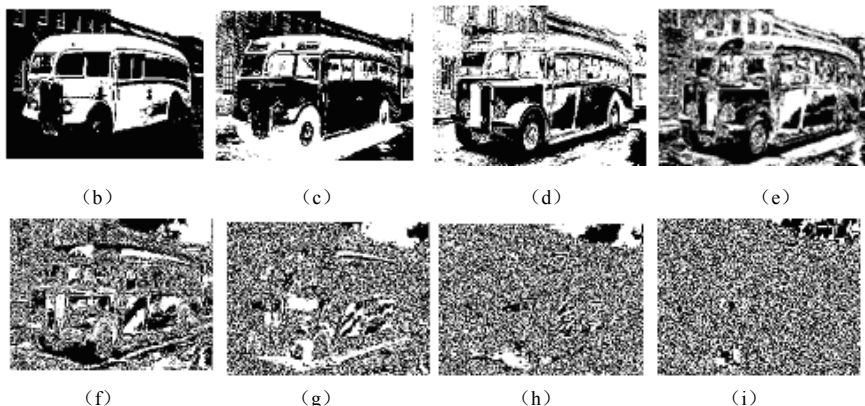


图 2.6 利用灰度码表示的位平面图

## 2) 位平面熵

由于图像在不同的位平面上有着不同的分布特性，如果只选择其中的一个位平面，则可能对图像不能进行很好的描述。同时可以看出，只有最高的几个位平面包含了视觉可见的有意义信息，体现了明显的图像结构特征；其他较低位平面的信息随机性很强，是很局部的小细节，只增加图像的亮度信息，没有提供任何的结构信息。因此，这里仅采用图像高位的 4 个位平面来提取图像特征。

为此，可以首先对图像通过位平面分解得到最高的 4 个位平面 ( $g_7, g_6, g_5, g_4$ )，然后转换成它们对应的灰度码表示的位平面 ( $G_7, G_6, G_5, G_4$ )，并计算每个位平面的信息熵。由于位平面为二值图像，所以位平面的信息熵 (Bit-plane Entropy, BE) 可简化为

$$BE = -p_1 \log_2 p_1 - p_0 \log_2 p_0 \quad (2-48)$$

其中， $p_1$  和  $p_0$  代表位平面中值为 1 及 0 的像素出现的概率。

位平面熵和图像的全局信息熵一样，具有旋转不变性、尺度不变性和平移不变性，对图像的几何形变具有很强的鲁棒性。同时，由于各个位平面实际上是提取了原图像中各个像素的灰度值的某一位形成的，因此，避免了直方图中颜色量化所带来的问题，而且也包含了图像像素的空间分布信息，解决了熵相同而图像不同的问题。熵矢量的计算只与每个位平面内图像像素点的数目有关，克服了原图像相邻像素点之间的相关性，忽略了图像的细微变化，即整体上求大同，局部上存小异。位平面熵的计算复杂度较低且维数较少，可以满足图像检索中存储容量和检索速度的要求。

### 3) 增强位平面熵<sup>[48]</sup>

#### (1) 改进的位平面熵

根据熵的对称性可知, 矢量各分量的次序任意改变时, 熵值不变, 熵函数的取值只与概率分布有关。如图 2.7 所示, (a)、(b) 表示两幅视觉效果完全不同的图像, (c)、(d) 为 (a)、(b) 的第 7 个位平面示意图, (e)、(f) 为 (c)、(d) 的统计直方图。可以看出, 它们的直方图差别很大, 但根据熵的对称性, 这两个直方图具有相似的信息熵。这样, 采样位平面熵就较难区分两个位平面。而且, 随着图像库中图像数量的增多, 出现这种情况的概率会越来越大, 也就是说, 随着图像库图像数量的增多, 位平面熵的检索准确度和检索率都会下降。

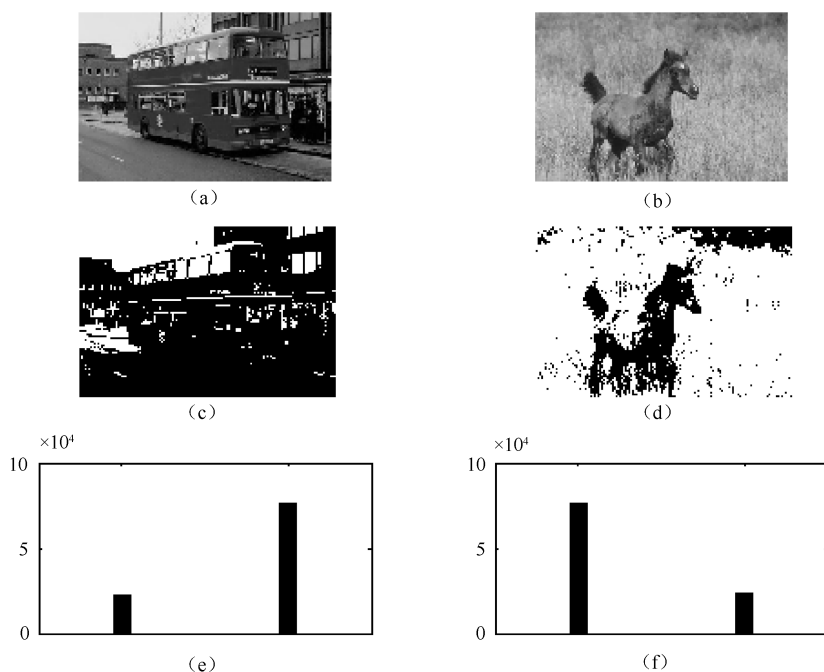


图 2.7 (a)、(b) 为两幅原始图像, (c)、(d) 为 (a) 及 (b) 的位平面 7, (e)、(f) 为 (c) 及 (d) 的直方图

为此, 我们给出了改进的位平面熵 (Enhanced Bit-plane Entropy, EBE), 其定义为

$$EBE = \begin{cases} BE, & p_1 > p_0 \\ -BE, & \text{其他} \end{cases} \quad (2-49)$$

这样, 对图 2.7 所示的位平面 7 来说, 它们的熵就有正负之分, 因此可以很容易进行区分。

但上述定义还存在一个问题。如图 2.8 所示, (a)、(b) 为两幅视觉效果相似的图

像, (c)、(d) 为 (a)、(b) 的第 7 个位平面示意图, (e)、(f) 为 (c)、(d) 的统计直方图。可以看出, 它们具有相似的直方图, 而且  $p_1$  和  $p_0$  的值也比较接近。但很明显, 对于 (e),  $p_1 > p_0$ , 对于 (f)  $p_1 < p_0$ 。在这种情况下, 如果还采用式 (2-49) 计算位平面熵, 它们的熵仍是一正一负, 这将导致检索错误。考虑到该问题, 我们对 EBE 作了进一步的定义, 为

$$EBE = \begin{cases} BE, & p_1 > p_0 \text{ 且 } p_1/p_0 < \alpha \\ -BE, & \text{其他} \end{cases} \quad (2-50)$$

其中,  $\alpha$  是事先设定的阈值。

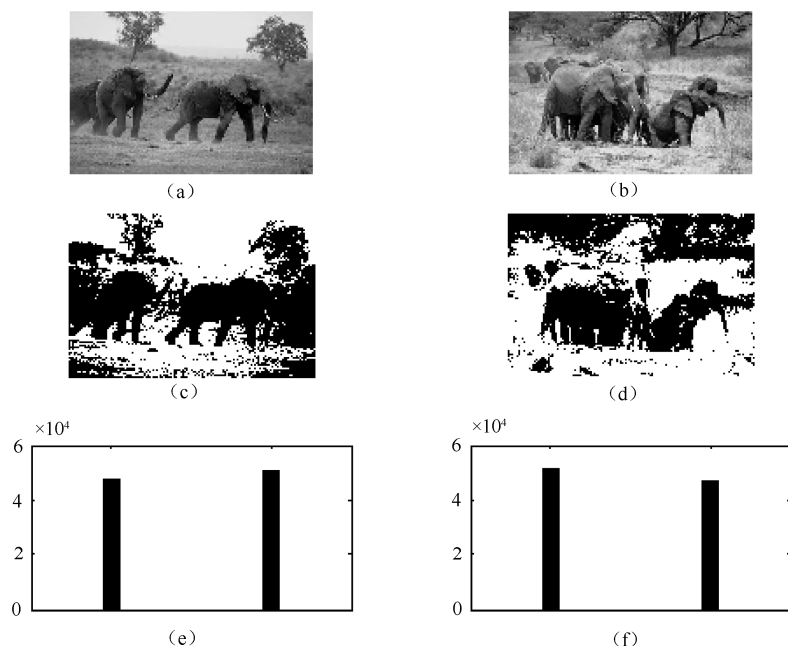


图 2.8 (a)、(b) 为两幅原始图像, (c)、(d) 为 (a) 及 (b) 的位平面 7, (e)、(f) 为 (c) 及 (d) 的直方图

## (2) 位平面空间分布熵

如图 2.9 所示, (a)、(b) 为两幅视觉效果完全不同的图像, (c)、(d) 为 (a)、(b) 的第 7 个位平面示意图, (e)、(f) 为 (c)、(d) 的统计直方图。可以看出, 两幅图像的位平面 7 具有十分相似的直方图, 它们具有相近的位平面熵。但从图中可明显看出, 虽然它们的直方图相似, 但具有不同值的像素 (黑、白两类) 的分布差别却很大, 也就是说它们具有不同的空间特征。因此, 空间特征也是影响检索准确度的重要因素。这里, 我们采用位平面空间分布熵来描述位平面的这种空间特征, 位平面空间分布熵的计算方法与前面介绍的颜色空间分布熵的计算方法相似, 这里不再详述。

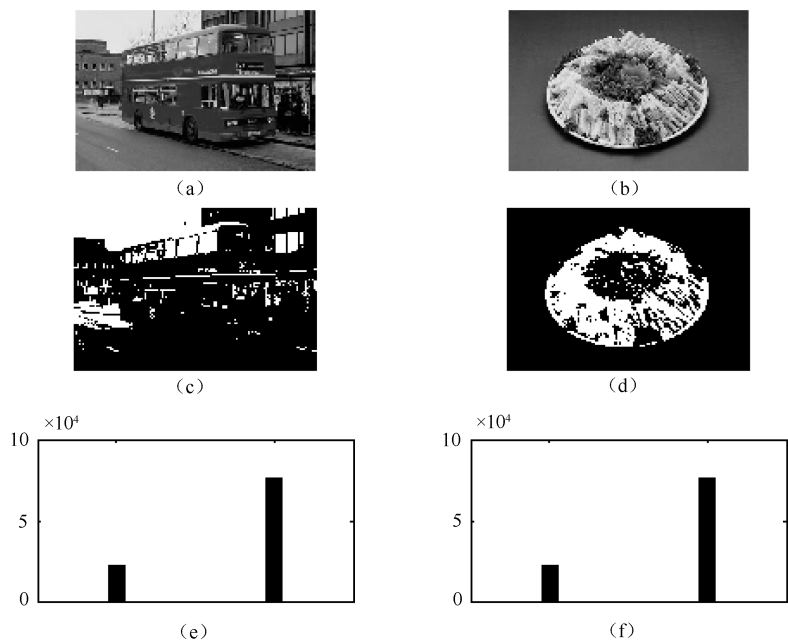


图 2.9 (a)、(b) 为两幅原始图像, (c)、(d) 为 (a) 及 (b) 的位平面 7, (e)、(f) 为 (c) 及 (d) 的直方图

## 7. 显著点颜色特征

显著点作为一种重要的图像视觉特征, 广泛用于三维解释、运动估计及图像匹配等方面。近几年来, 显著点被广泛用于图像检索<sup>[28]</sup>。下面主要介绍在图像的块逆概率差 (Block Difference of Inverse Probabilities, BDIP)<sup>[49]</sup>模型的基础上, 得到原图像的 BDIP 图像, 然后根据 BDIP 图像中像素的分布特点来提取图像的显著点, 然后以它们为线索, 把图像的形状特征和空间颜色分布特征有机地结合起来进行检索。

### 1) 块逆概率差 (BDIP) 模型及 BDIP 图像的提取

图像的块逆概率差 (BDIP) 定义为

$$\text{BDIP} = M^2 - \frac{\sum_{(i,j) \in \mathbf{B}} I(i,j)}{\max_{(i,j) \in \mathbf{B}} I(i,j)} \quad (2-51)$$

其中,  $I(i,j)$  表示像素点  $(i,j)$  处的灰度值;  $\mathbf{B}$  表示大小为  $M \times M$  的图像块。

用图像块的 BDIP 值表示的图像, 我们称之为原图像的 BDIP 图像。在利用这个模型提取图像的显著点时, 由于 BDIP 图像中像素值的分布比较分散, 因此需要对其进行归一化处理。归一化的 BDIP 值定义为

$$\text{BDIP} = \frac{M^2 - \frac{\sum_{(i,j) \in \mathbf{B}} I(i,j)}{\max_{(i,j) \in \mathbf{B}} I(i,j)}}{M^2} \quad (2-52)$$

对一幅真彩色图像来说，在提取 BDIP 图像时要考虑选择一个合适的颜色空间，提取的 BDIP 图像才能更好地刻画原始图像的内容特征。这里采用 HSV 颜色空间，并采用黄元元提出的量化方法<sup>[3]</sup>，将图像的颜色量化为 36 柄。量化过程如下。

$$H = \begin{cases} 0, & H \in [0^\circ, 60^\circ) \\ 1, & H \in [60^\circ, 120^\circ) \\ 2, & H \in [120^\circ, 180^\circ) \\ 3, & H \in [180^\circ, 240^\circ) \\ 4, & H \in [240^\circ, 300^\circ) \\ 5, & H \in [300^\circ, 360^\circ) \end{cases} \quad S = \begin{cases} 0, & S \in [0, 0.25) \\ 1, & S \in [0.25, 1] \end{cases} \quad V = \begin{cases} 0, & V \in [0, 0.3) \\ 1, & V \in [0.3, 0.8) \\ 2, & V \in [0.8, 1.0] \end{cases} \quad (2-53)$$

根据式 (2-53) 的划分，我们可以将很多虽然深浅不同但在视觉上仍属于同一类的颜色量化在同一区间内，使量化结果更加符合人类的视觉感受。

图 2.10 给出了用 BDIP 值表示的原图像，即图像的 BDIP 图像，子图像块的大小取为  $3 \times 3$ 。在图 2.10 中，(a) 是原始图像，图像大小为  $384 \times 256$ ；(b) 是原图像的 BDIP 图像，图像大小为  $128 \times 85$ 。由图中可以看出，对于灰度变化明显的块，其 BDIP 值也会很大。而且 BDIP 图像中的像素点几乎都位于图像中能引起注意的视觉焦点位置，因此利用 BDIP 图像可以很好地表示图像的基本特征。

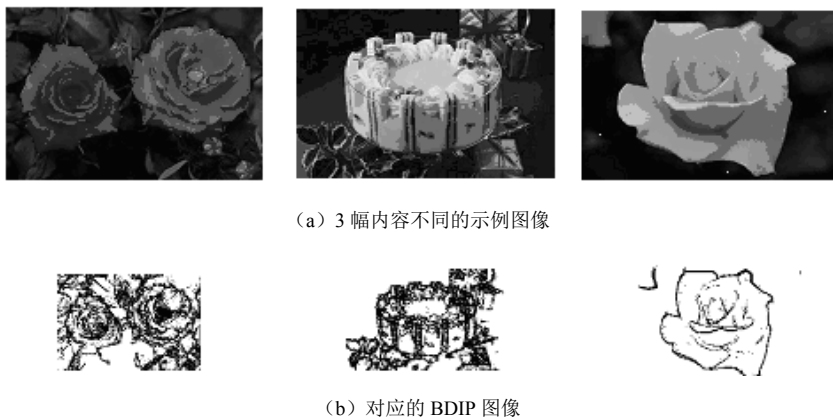


图 2.10 示例图像及其对应的 BDIP 图像

## 2) 显著点提取算法

对一幅原图像变换后得到的 BDIP 图像来说，其中的像素点能完整地表示出图像

的内容特征,考虑到这点,在算法中将 BDIP 图像中灰度值不为 0 的像素点先作为显著点的初选。这些显著点并不都是重要的,要选取一些突出的足够引起视觉注意的显著点。另外,还应该考虑到,所选取的显著点个数对图像的检索效果有很大的影响,显著点选取的较少,体现不出图像的特征,若选取的过多,会增加图像特征提取的计算量。由于图像块中的灰度变化程度不仅可以用图像的 BDIP 值来反映,还可用块内灰度值的方差来体现。方差的大小和 BDIP 值的大小是一致的。BDIP 值大,块内的方差也大,反之亦然。因此结合块内灰度值的方差  $\sigma$ , 给出一个有一定自适应性的显著点选择条件,为

$$\sigma \times V_{\text{BDIP}} \geq \lambda \mu_T \quad (2-54)$$

其中,  $V_{\text{BDIP}}$  为图像块的 BDIP 值;  $\lambda$  为实验中需设定的参数;  $\mu_T$  是整幅图像中图像块 BDIP 值的均值。这样,一幅图像中显著点的选择就可以根据自身图像块的方差、BDIP 值及整幅图像中图像块 BDIP 值的均值决定。

图 2.11 给出了几幅图像的 BDIP 图像及提取的显著点。从图中可以看出,式(2-54)中的条件倾向于选取块内方差较大且 BDIP 值较大的显著点,而且采用该算法不仅可以检测出角点,还能检测出平滑边缘上的点。图中显著点都集中在图像中视觉焦点上,而且显著点构成的轮廓可以清晰地描述图中物体的形状。边缘点和角点都是引起人视觉注意的地方,在这些地方提取图像特征,有利于图像检索结果和人的视觉保持一致。

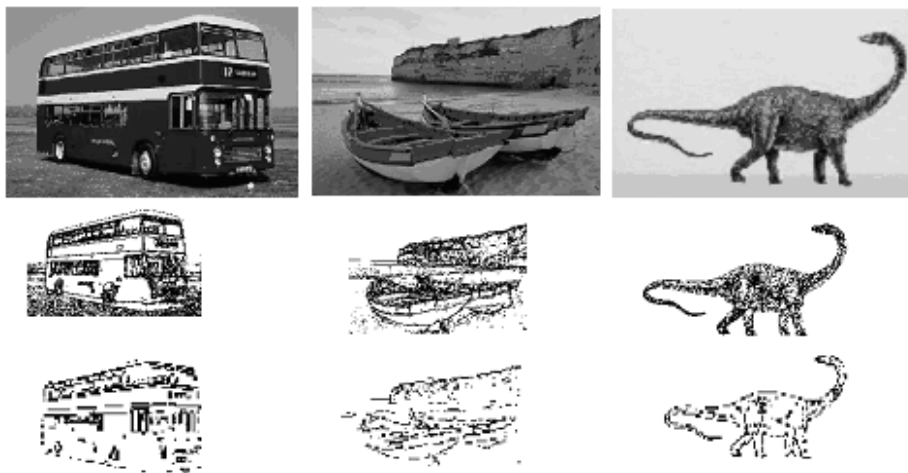


图 2.11 示例图像及相应的 BDIP 图像和显著点提取图像

### 3) 基于显著点的特征提取

由于显著点往往分布在图像中视觉上最感兴趣的地方,蕴含了丰富的颜色细节,所以利用显著点局部区域的颜色特征来描述图像具有合理性。每个显著点所对应的 BDIP 值在一定程度上可以体现其与周围点的灰度对比度, BDIP 值越大,显著点对

应的灰度值也越大。因此，可以结合 BDIP 值的分布特征提取显著点的颜色空间分布特征。

为了提取颜色的空间分布信息，这里对显著点的 BDIP 值的分布特征进行了统计，所选取的图像被划分为  $3 \times 3$  大小的子块，实验结果表明，大部分显著点的值在  $[0, 0.1]$ ，少量的值分布在  $[0.1, 0.4]$ ，只有极少一部分取值在  $[0.4, 1]$ 。

根据上述显著点的分布，可以将其分成 3 种状态：密集态、常态和稀疏态。通过提取这 3 种状态下显著点颜色分布的一阶矩和二阶矩来描述颜色及其空间分布信息。定义如下。

$$\begin{aligned}\mu_p &= \frac{1}{N_p} \sum_{V_{\text{BDIP}} \in p} V_{\text{BDIP}} \\ \sigma_p &= \sqrt{\frac{1}{N_p} \sum_{V_{\text{BDIP}} \in p} (V_{\text{BDIP}} - \mu_p)^2}\end{aligned}\tag{2-55}$$

其中， $p=0,1,2$ ，分别表示显著点的 3 种状态； $N_p$  表示属于  $p$  状态下显著点的数目。

## 2.2 形状特征的提取与表达

### 2.2.1 概述

相对于颜色或纹理等低层特征而言，形状特征属于图像的中间层特征，它作为刻画图像中物体和区域特点的重要特征，是描述高层视觉特征（如目标、对象）的重要手段，而目标、对象对获取图像语义尤为重要。对利用形状特征的图像检索，人们已提出了许多不同的方法，有关基于形状特征的检索算法可参阅相关综述性文献[50]、[51]、[52]、[53]等。

按照形状表达的形式划分，形状的描述符可分为两大类：第一类是表达形状的目标区域边界轮廓的像素集合，称之为基于轮廓的形状描述符（contour-based）；第二类是表达形状的目标区域内的所有的像素集合，称之为基于区域的形状描述符（region-based）。详细的形状特征描述划分方法如图 2.12 所示<sup>[51]</sup>。



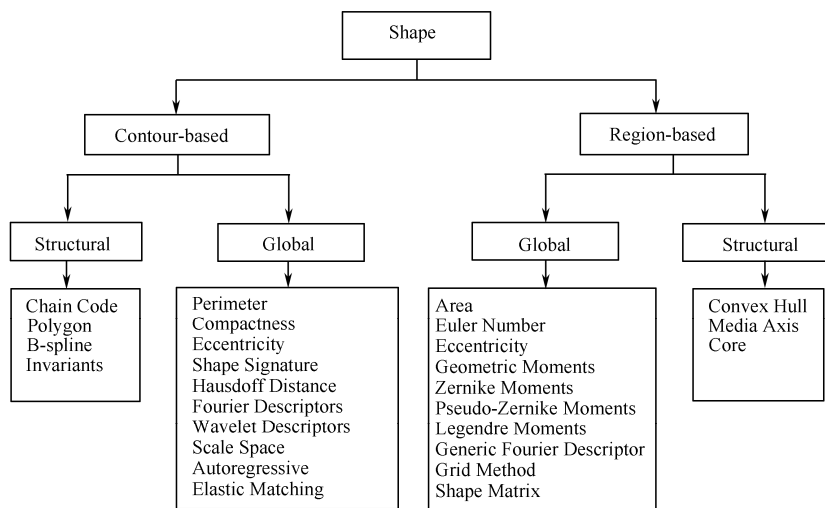


图 2.12 形状特征描述技术分类

## 2.2.2 基于轮廓的描述方法

基于轮廓的描述方法仅仅提取形状的轮廓信息，这类描述方法一般有两种形式：连续型（即全局型）和离散型（即结构型）。连续型的描述方法不对轮廓进行分段处理，往往是从全局轮廓抽取特征向量；离散型的描述方法往往首先将轮廓划分为很多片段，然后提取相应的特征。基于轮廓的常用方法有 Freeman 链码、Shape Contexts、曲率尺度空间描述符、BAS（Beam Angle Statistics）、TAR（Triangel-Area Representation）、傅里叶描述符、小波描述符与边界矩等<sup>[1]</sup>。为了比较不同轮廓描述符的性能，文献[54]对 15 种轮廓描述符的检索效果进行了比较，比较结果可参见 <http://give-lab.cs.uu.nl/sidestep/>。

本小节主要介绍 7 种基于轮廓的形状描述方法：简单几何参数描述符、傅里叶形状描述符、曲率尺度空间描述符、小波描述符、方向链码、基于角点的描述方法与基于矩的描述方法。

### 1. 简单几何参数描述符

简单的轮廓几何参数描述符主要包括以下几种。

#### 1) 边界长度

边界长度是一种简单的边界全局特征，它是边界所包围区域的轮廓的周长。边界长度一般常用 4-方向连通边界或 8-方向连通边界表示，并相应地得到一个近似长度。

## 2) 边界直径

边界直径是边界上相隔最远的两点之间的距离，即这两点之间的连线段长度。有时这条线段也称为边界的主轴或长轴（与此垂直且与边界的两个交点间的线段最长的也叫边界的短轴）。它的长度和取向对描述边界都很有用。

## 3) 曲率

曲率是斜率的改变率，它描述了边界上各点沿边界方向变化的情况。对离散的轮廓图来说，曲率往往会受到噪声及边缘细节的影响。

## 4) 形状数

形状数是基于链码的一种边界形状描述符，形状数提供了一种有效的形状度量方法，它的每个阶是唯一的，不随边界的旋转和尺度的变化而改变。

## 2. 傅里叶形状描述符

傅里叶形状描述符 (Fourier Shape Descriptors) 是一种广泛应用的形状描述符，其基本思想是用物体边界的傅里叶变换作为其形状描述<sup>[51]</sup>。假设一个二维物体的轮廓是由一系列坐标为  $(x_s, y_s)$  的像素组成的，其中  $0 \leq s \leq N-1$ ，而  $N$  是轮廓上像素的总数。从这些边界点的坐标中可以推导出 4 种形状表达，分别是曲率函数 (curvature function)、质心距离 (centroid distance)、复坐标函数 (complex coordinates function) 和弦长函数 (chord function)。下面给出文献[55]对傅里叶描述符的描述。

轮廓线上某一点的曲率定义为轮廓线的切向角度相对于弧长的变化率。曲率函数  $K(s)$  可以表示为

$$K(s) = \frac{d}{ds} \theta(s) \quad (2-56)$$

其中,  $\theta(s)$  是轮廓线的切向角度，定义为

$$\left. \begin{aligned} \theta(s) &= \argtan\left(\frac{y'_s}{x'_s}\right) \\ y'_s &= \frac{dy_s}{ds} \\ x'_s &= \frac{dx_s}{ds} \end{aligned} \right\} \quad (2-57)$$

质心距离定义为从物体边界点到物体中心  $(x_c, y_c)$  的距离，即

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (2-58)$$

复坐标函数是用复数所表示的像素坐标，即

$$Z(s) = (x_s - x_c) + j(y_s - y_c) \quad (2-59)$$

对这种复坐标函数的傅里叶变换会产生一系列复数系数。这些系数在频率上表示

了物体形状，其中，低频分量表示形状的宏观属性，高频分量表达了形状的细节特征。形状描述符可以从这些变换参数中得出。为了保持旋转无关性，可以仅仅保留参数的大小信息，而省去相位信息。缩放的无关性是通过将参数的大小除以直流分量（即第一个非零参数）的大小来保证的。另外，变换无关性是基于轮廓的形状表示所固有的特点。

对于曲率函数和质心距离，我们只考虑正频率的坐标轴，因为这时函数的傅里叶变换是对称的，即有  $|F_{-i}| = |F_i|$ 。基于曲率函数的形状描述符表示为

$$f_K = [|F_1|, |F_2|, \dots, |F_{M/2}|] \quad (2-60)$$

其中， $F_i$  表示傅里叶变换参数的第  $i$  个分量。类似地，由质心距离所导出的形状描述符为

$$f_R = \left[ \frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{M/2}|}{|F_0|} \right] \quad (2-61)$$

对于复坐标函数，正频率分量和负频率分量被同时采用。由于直流分量与形状所处的位置有关而被省去。同时，第一个非零的频率分量被用来对其他变换参数进行标准化。由复坐标函数所导出的形状描述符为

$$f_Z = \left[ \frac{|F_{-(M/2-1)}|}{|F_1|}, \dots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \dots, \frac{|F_{M/2}|}{|F_1|} \right] \quad (2-62)$$

为保证数据库中所有物体的形状特征都有相同的长度，在实施傅里叶变换之前需要将所有边界点的数目统一，这样就可以采用快速傅里叶变换来提高算法效率。

### 3. 曲率尺度空间描述符

曲率尺度空间描述符（Curvature Scale Space Descriptor, CSSD）由 Mokhtarian 等引入并成功应用于形状特征的描述及检索中<sup>[56]</sup>。目前，该描述符成为 MPEG-7 标准中轮廓描述的一种方法。该方法的基本思想是基于人们在认知物体时，倾向于将物体的形状分解成凹和凸的部分来比较这一特性。CSSD 是在多尺度下根据变形点将轮廓“分解”成凹和凸部分，这些变形点定义为轮廓曲线上曲率为零的点。

给出物体的闭合轮廓曲线，对曲线的两个直角坐标  $x$  及  $y$  进行参数化，采用自然参数即曲线的弧长为参数，以任意一点为起点，顺时针跟踪轮廓，并归一化使得弧长参数  $l \in [0, 1]$ ，在曲线起始点处  $l = 0$ ，终点处  $l = 1$ ，轮廓曲线表示为  $C = \{x(l), y(l) \mid l \in [0, 1]\}$ 。轮廓若为闭合，则起点和终点是重合的，且  $x(l)$  和  $y(l)$  为以 1 为周期的周期函数。曲率计算为

$$k(l) = \frac{x'(l)y''(l) - x''(l)y'(l)}{\left\{ [x'(l)]^2 + [y'(l)]^2 \right\}^{\frac{3}{2}}} \quad (2-63)$$

其中， $x'(l)$ 、 $y'(l)$ 、 $y''(l)$  及  $x''(l)$  分别代表一阶及二阶导数。

为了获得不同分辨尺度下轮廓的曲率，需要对轮廓进行演化，用核宽为 $\sigma$ 的一维高斯函数 $g(l, \sigma)$ 进行卷积平滑，然后在平滑的基础上再次计算曲率。图 2.13 给出了轮廓在不同尺度下的曲率过零点的情况，图 2.14 给出了轮廓的 CSSD 示意图。

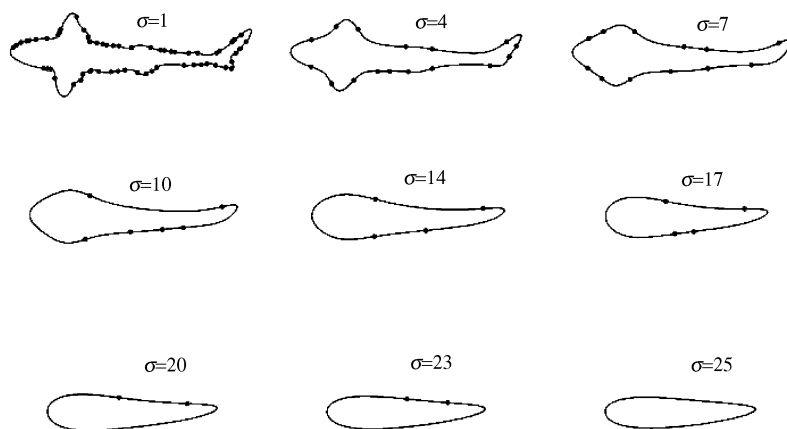


图 2.13 轮廓在不同尺度下的曲率过零点的情况

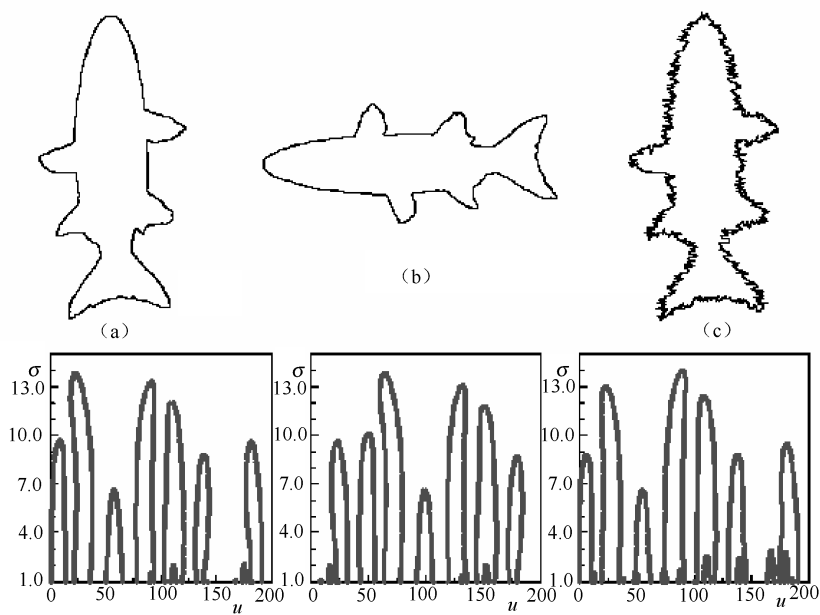


图 2.14 轮廓的 CSSD 示意图

#### 4. 小波描述符

利用小波变换对图像轮廓进行描述<sup>[57,58]</sup>，首先要定义小波函数族。小波函数族定义为

$$\psi_{mn}(t) = 2^{-m/2} \psi(2^{-m}t - n) \quad (2-64)$$

假设图像的轮廓函数为  $f(t)$ ，它的小波变换系数为

$$c_{mn}(t) = \int_{-\infty}^{\infty} f(t) \psi_{mn}(t) dt \quad (2-65)$$

利用小波系数可以重建  $f(t)$ ，重建过程如下。

$$f(t) = \sum_{m=m_0+1}^{\infty} \sum_{n=-\infty}^{\infty} c_{mn} \psi_{mn}(t) + \sum_{m=-\infty}^{m_0} \sum_{n=-\infty}^{\infty} c_{mn} \psi_{mn}(t) \quad (2-66)$$

其中， $m_0$  与截断系数时所需的精度相关。

假设尺度函数为  $S_{mn}(t) = 2^{-m/2} S(2^{-m}t - n)$ ，结合小波重建公式，则有

$$f(t) = \sum_{n=-\infty}^{\infty} c_{mn} S_{mn}(t) + \sum_{m=-\infty}^{m_0} \sum_{n=-\infty}^{\infty} c_{mn} S_{mn}(t) \quad (2-67)$$

如果  $c_{mn} S_{mn}(t)$  称为尺度系数，则  $c_{mn} \psi_{mn}(t)$  称为小波系数，所有的小波系数组成轮廓相对应的小波轮廓描述符。

小波轮廓描述符在粗尺度上给出了形状的全局信息，在细尺度上给出了局部信息。由于小波变换提供了多分辨率表示，因此识别过程可以根据输入图像或目标而灵活调整。但小波描述符依赖于目标曲线的起始点，也就是说，同一目标的两条不同采样曲线的小波表示可能因为起始点不同而有很大的差异。

#### 5. 方向链码

##### 1) 链码

链码是对图像边界点的一种编码表示方法，其特点是利用一系列具有特定长度和方向的直线段相连来表示目标的边界。因为每个线段的长度固定而方向数目取为有限，所以只有边界的起点需用坐标表示，其余点都只需用连续方向来代表偏移量。常用的链码有4-方向链码和8-方向链码，其方向定义如图2.15所示。在此基础上，Freeman推广了原来的定义获得了广义链码，并且还利用链码来抽签关键点，从而生成一种相对于平移、旋转、尺度不变的表示方法，他还总结了链码的各种算法<sup>[59]</sup>。由于链码表示简单且所需的存储空间较小，因此链码广泛地应用到形状编码及模式识别中<sup>[60-62]</sup>。同时，在链码定义的启发下，还出现了其他一些定义链码的方法，如夹角链码<sup>[63]</sup>及可变夹角链码<sup>[64]</sup>等。文献[65]、[66]对多种链码描述轮廓的性能进行了比较分析。

针对方向链码，起点的变化、尺度的变化、目标的旋转等都会引起链码串的变化，而在对形状进行描述时，关键就是要使所选择的描述符满足尺度、旋转和平移不变性。

因此，往往需要进行起点归一化及旋转归一化。

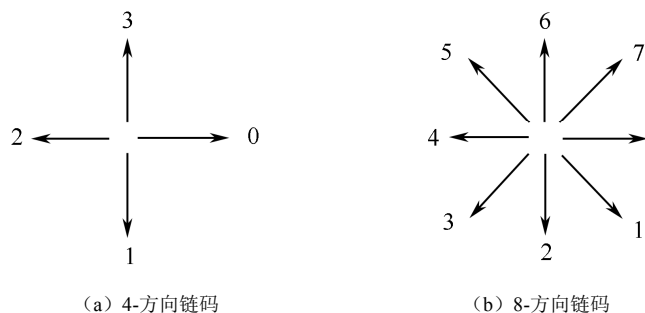


图 2.15 4-方向链码及 8-方向链码

作为一种有效的轮廓描述方法，链码广泛地应用到形状的识别和编码技术中。但其在图像检索中的应用相对还较少，常用的方法有链码直方图<sup>[67]</sup>、最小和统计方向码<sup>[68]</sup>及我们提出的相关改进方法<sup>[69,70]</sup>等。

## 2) 链码直方图 (Chain Code Histogram, CCH)

链码直方图的定义为

$$h_i = \frac{n_i}{N} \quad (2-68)$$

其中， $n_i$  表示链码串中  $i$  方向链码的个数； $N$  表示链码串中所有链码的数目。可以看出，链码直方图反映了不同方向链码在链码串中出现的概率，是一种统计特征，且与起点的选择无关，并且具有尺度及平移不变性，但不具备旋转不变性。

## 3) 最小和统计方向码 (MSSDC)

首先给出统计方向码的定义。沿顺时针方向，对象轮廓方向码定义为

$$\mathbf{X} = (x_0 \ x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7) \quad (2-69)$$

其中， $x_i$  是边缘上具有方向码  $i$  的像素数量，故统计方向码与起始点无关，并且天然具备了对图像平移的不变性。构造方向码向量，表示链码的 8 个方向，即

$$\mathbf{D} = [0 \ 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7]^T \quad (2-70)$$

设对象轮廓的初始方向码表示为  $\mathbf{X}_0 = (x_0 \ x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7)$  (顺时针方向)，当图像顺时针依次旋转  $45^\circ$ ，统计方向码的变化有如下规律。

$$\mathbf{X}_i = (x_{(i+0) \oplus 8} \ x_{(i+1) \oplus 8} \ x_{(i+2) \oplus 8} \ x_{(i+3) \oplus 8} \ x_{(i+4) \oplus 8} \ x_{(i+5) \oplus 8} \ x_{(i+6) \oplus 8} \ x_{(i+7) \oplus 8}) \quad (2-71)$$

其中， $\oplus 8$  是模 8 运算，因为方向码的变化是以 8 次为一个周期。可见，由于图像旋转，会导致同一轮廓的方向码发生变化，即缺乏旋转不变性。为了解决该问题，文献<sup>[68]</sup>给出了最小和统计方向码的定义，即

$$D_{\min} = \min \{X_i \cdot D | i = 0, 1, \dots, 7\} \quad (2-72)$$

最小和统计方向码通过最小和的约束，实际上在 8 个可能的方向中唯一确定其中某一视角描述对象形状。这时，统计方向码才可以被用来作为唯一索引。同时，文中还规定，如果有多个统计方向码满足上述条件，则选择沿顺时针方向遇到的第一个方向码为最终表示，因此最小和统计方向码在一定程度上解决了旋转不变性的问题。同时，文章也证明最小和统计方向码具有成比例的特点，这为消除尺度不变性创造了条件。

#### 4) 改进链码特征

上述 CCH 和 MSSDC 方法存在一个共同的问题，都没有考虑方向链码的空间分布特征。设 {000007666665444443222221} 及 {666654443220001222007644} 分别表示两个边缘轮廓的链码串，它们表示的形状如图 2.16 (a) 和 (b) 所示。按照上述方法，它们具有相同的 CCH 及 MSSDC，可是它们所表示的边缘形状却大不相同，主要原因就是上述方法没有考虑方向链码的空间分布特征。

从上面的分析可以看出，链码的空间分布特征也是边缘形状的一种重要特征，几种新的空间特征描述方法包括链码分布矢量 (Chain Code Distribution Vector, CCDV)、链码相关矢量 (Chain Code Coherence Vector, CCCV)、链码熵 (Chain Code Entropy, CCE)、链码空间分布熵 (Chain Code Spatial Distribution Entropy, CCSDE)、链码相关熵 (Chain Code Relativity Entropy, CCRE) [69, 70]。

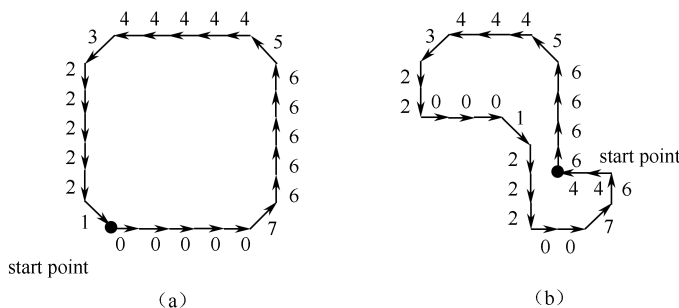


图 2.16 示例轮廓链码表示

##### (1) 链码分布矢量 (CCDV)

**定义 2-1** 设  $S$  代表任一轮廓的链码串， $s_i$  ( $s_i \subset S$ ) 代表链码串中相连的、方向为  $i$  的链码序列，定义链码序列  $s_i$  在  $S$  中出现的序号为该链码串的距离，同时规定链码串中起始链码序列的距离为 0。

根据该定义，上述两个链码串的链码距离如图 2.17 所示。针对  $i$  方向链码，其距离分布如图 2.18 所示。其中， $S$  表示链码串的链码； $j$  表示  $i$  方向链码出现的位置（从 0 开始计）， $m_j$  表示  $i$  方向链码在  $j$  位置连续出现的次数， $d_j$  表示  $j$  位置  $i$  方向链码序列的距离。

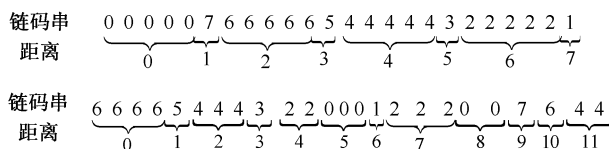


图 2.17 链码距离计算示例

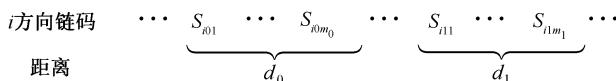


图 2.18  $i$  方向链码距离分布

定义了链码距离后，我们可以根据链码距离的变化来反映不同方向链码的空间分布特征，具体地说，我们采用链码距离的均方差来表示，其定义为

$$\sigma_i = \sqrt{\frac{m_0(d_0 - \mu_i)^2 + m_1(d_1 - \mu_i)^2 + \cdots + m_j(d_j - \mu_i)^2 + \cdots}{m_0 + m_1 + \cdots + m_j + \cdots}} \quad (2-73)$$

其中， $\mu_i = \frac{m_0 d_0 + m_1 d_1 + \cdots + m_j d_j + \cdots}{m_0 + m_1 + \cdots + m_j + \cdots}$ 。

这里我们在 MSSDC 的基础上来获取链码的空间分布特征，即首先计算 MSSDC，然后再获取其空间分布特征。由于 MSSDC 具有旋转和平移不变性，因此  $\sigma_i$  也具备相应的性质。从链码距离的定义可以看出，链码距离与起点的选择有关，因此  $\sigma_i$  也与起点的选择相关。

当某一对像的尺度发生了变化（如放大或缩小  $\lambda$  倍）， $\sigma_i$  具有尺度不变性。由于尺度的变化仅仅引起链码串长度的变化，从链码距离的定义可以看出，尺度的变化对链码距离没有影响，因此有

$$\mu_i^{(\lambda)} = \frac{(\lambda m_0) d_0 + (\lambda m_1) d_1 + \cdots + (\lambda m_j) d_j + \cdots}{\lambda m_0 + \lambda m_1 + \cdots + \lambda m_j + \cdots} = \mu_i \quad (2-74)$$

$$\begin{aligned} \sigma_i^{(\lambda)} &= \sqrt{\frac{(\lambda m_0)(d_0 - \mu_i^{(\lambda)})^2 + (\lambda m_1)(d_1 - \mu_i^{(\lambda)})^2 + \cdots + (\lambda m_j)(d_j - \mu_i^{(\lambda)})^2 + \cdots}{\lambda m_0 + \lambda m_1 + \cdots + \lambda m_j + \cdots}} \\ &= \sqrt{\frac{m_0(d_0 - \mu_i)^2 + m_1(d_1 - \mu_i)^2 + \cdots + m_j(d_j - \mu_i)^2 + \cdots}{m_0 + m_1 + \cdots + m_j + \cdots}} \\ &= \sigma_i \end{aligned} \quad (2-75)$$

可以看出，尺度的变化并没有对  $\sigma_i$  产生影响。

这里，我们就采用  $\sigma_i$  作为链码的分布特征，结合链码直方图，我们给出了 CCDV 的定义如下。

$$\langle (h_0, \sigma_0), (h_1, \sigma_1), \cdots, (h_i, \sigma_i), \cdots, (h_{n-1}, \sigma_{n-1}) \rangle \quad (2-76)$$



其中,  $n$  表示链码的方向数。

针对上面的两个链码串 {000007666665444443222221} 及 {666654443220001222007644}, 其 CCDV 可分别表示为:  $\langle (0.21, 0), (0.04, 0), (0.21, 0), (0.04, 0), (0.21, 0), (0.04, 0), (0.21, 0), (0.04, 0) \rangle$  及  $\langle (0.21, 1.64), (0.04, 0), (0.21, 1.64), (0.04, 0), (0.21, 4.93), (0.04, 0), (0.21, 4.47), (0.04, 0) \rangle$ 。

可以看出, 虽然两个链码串具有相同的统计特征, 但是其空间分布特征却有较大的区别, 因此通过其空间分布特征可较好地地区分两个链码串。

### (2) 链码相关矢量 (CCCV)

针对该方法, 我们将链码串中的链码划分为聚合链码和非聚合链码两种类型, 其定义如下。

**定义 2-2** 在链码串中, 针对  $i$  方向链码, 若  $m_j \geq \tau$  ( $\tau$  为给定的阈值), 则定义链码序列  $S_{ij0}, S_{ij1}, \dots, S_{ijm_j}$  为聚合链码, 否则为非聚合链码。

设  $\alpha_i$  表示  $i$  方向聚合链码的统计数目,  $\beta_i$  表示  $i$  方向非聚合链码的统计数目, 则 CCCV 可定义为

$$\langle (\alpha_0, \beta_0), (\alpha_1, \beta_1), \dots, (\alpha_i, \beta_i), \dots, (\alpha_{n-1}, \beta_{n-1}) \rangle \quad (2-77)$$

其中,  $n$  表示链码的方向数。

例如, 针对上文链码串 {000007666665444443222221} 及 {666654443220001222007644}, 如果我们选择  $\tau = 4$ , 则其 CCCV 可分别表示为:  $\langle (5, 0), (0, 1), (5, 0), (0, 1), (5, 0), (0, 1), (5, 0), (0, 1) \rangle$  及  $\langle (4, 1), (0, 1), (0, 5), (0, 1), (0, 5), (0, 5), (0, 1), (0, 1) \rangle$ 。

可以看出, 虽然它们具有相同的统计特征, 但其 CCCV 也存在较大区别。

这里, 我们仍在 MSSDC 的基础上提取 CCCV, 因此 CCCV 也具有平移及旋转不变性。很明显, CCCV 与不同方向链码的数目有关, 因此不具备尺度不变性。为了解决这个问题, 我们又对该 CCCV 进行了归一化处理。

设尺度变化为  $\lambda$ , 如果我们取阈值  $\tau' = \lambda\tau$ , 从而不同尺度下的 CCCV 可表示为  $\langle (\lambda\alpha_0, \lambda\beta_0), (\lambda\alpha_1, \lambda\beta_1), \dots, (\lambda\alpha_{n-1}, \lambda\beta_{n-1}) \rangle$ , 设  $L_1 = \sum_{i=0}^{n-1} \alpha_i$ ,  $L_2 = \sum_{i=0}^{n-1} \beta_i$ , 经过归一化处理可表示为

$$\langle (\frac{\alpha_0}{L_1}, \frac{\beta_0}{L_2}), (\frac{\alpha_1}{L_1}, \frac{\beta_1}{L_2}), \dots, (\frac{\alpha_i}{L_1}, \frac{\beta_i}{L_2}), \dots, (\frac{\alpha_{n-1}}{L_1}, \frac{\beta_{n-1}}{L_2}) \rangle \quad (2-78)$$

显然, 通过上述处理, CCCV 具有尺度不变性。

### (3) 链码空间分布熵 (CCSDE)

链码空间分布熵的计算方法同 2.1.4 小节中的颜色空间分布熵的计算方法。

### (4) 链码相关熵 (CCRE)

针对如图 2.19 所示的轮廓, 其 4-方向链码可表示为 {10103322}, 从图中可以看出, 该轮廓的主要变化体现在链码  $1 \rightarrow 0$ ,  $0 \rightarrow 1$ ,  $1 \rightarrow 0$ ,  $0 \rightarrow 3$ ,  $3 \rightarrow 2$ ,  $2 \rightarrow 1$  链码方向的转

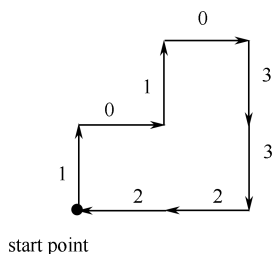


图 2.19 轮廓示例

换, 因此链码间的方向变化也是反映图像轮廓变化的一个主要特征, 而前面所述的方法都没有考虑这种特征。

如果我们将链码方向看作不同的状态，那么链码串就可以看作状态链，从而链码方向的变化可以通过状态的变化来体现。很明显，采用马尔可夫链可以很好地解决该问题，马尔可夫链的状态转移矩阵就反映了状态间的变化，这里我们采用一步转移概率矩阵来描述。为了获取更有效的特征表示，我们对马尔可夫链的一步转移概率矩阵进行了修改，新定义的状态转移概率矩阵如下。

$$\mathbf{P} = \begin{bmatrix} 0 & p_{01} & \cdots & p_{0(n-1)} \\ p_{10} & 0 & \cdots & p_{1(n-1)} \\ \vdots & \vdots & \vdots & \vdots \\ p_{(n-1)1} & p_{(n-1)2} & \cdots & 0 \end{bmatrix} \quad (2-79)$$

其中,  $p_{ij} = \begin{cases} \frac{k_{ij}}{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} k_{ij}}, & i \neq j \\ 0, & i = j \end{cases}$ ,  $k_{ij}$  表示从状态  $i$  转移到状态  $j$  的转移次数,  $\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} p_{ij} = 1$ 。

在原马尔可夫链的状态转移概率矩阵中,  $p_{ij} = \frac{k_{ij}}{\sum_{i=0}^{n-1} k_{ij}}$  且  $\sum_{j=0}^{n-1} p_{ij} = 1$ 。

在此基础上, 我们给出了状态  $i$  的相关性直方图定义, 即

$$\mathbf{q}_i = (p_{i0}, p_{i1}, \dots, p_{ij}, \dots, p_{i(n-1)}, p_{0i}, p_{1i}, \dots, p_{ji}, \dots, p_{(n-1)i}), \quad i \neq j \quad (2-80)$$

该相关性直方图包含两个部分：一部分是状态  $i$  到其他状态的转移概率；另一部分是其他状态到状态  $i$  的转移概率。该相关性直方图可以简单地表示为

$$\mathbf{q}_i = (q_1, q_2, \cdots, q_k, \cdots, q_{n-2}) \quad (2-81)$$

其中,  $q_1 = p_{i0}, q_2 = p_{i1}, \dots, q_{2n-2} = p_{(n-1)i}$ 。可以看出, 如果状态数为  $n$ , 则相关性直方图的维数为  $2n-2$ , 结合信息熵的定义, 我们给出了  $i$  方向链码的相关熵定义, 即

$$\text{RE}_i = -\sum_{k=1}^{2n-2} q_k \log_2(q_k) \quad (2-82)$$

$RE_i$ 反映了状态*i*与其他状态的相关性,也即反映了该方向链码同其他方向链码的相关性, $RE_i$ 越大,表明相关性越强。结合链码直方图,新的特征向量可以描述为

$$\langle (h_0, RE_0), (h_1, RE_1), \dots, (h_i, RE_i), \dots, (h_{n-1}, RE_{n-1}) \rangle \quad (2-83)$$

## 6. 基于角点的描述方法

关于轮廓角点提取,目前的方法很多,下面主要介绍3种不同的角点提取方法,并在此基础上,讨论如何有效地提取角点特征并用于基于轮廓的图像检索。

### 1) 轮廓角点检测

#### (1) 基于内角的角点检测

针对轮廓角点, 我们可根据其内角大小简单判断<sup>[71]</sup>。轮廓首先被均匀采样, 设采样后的轮廓表示为  $\{s_1, s_2, \dots, s_i, \dots, s_n\}$ ,  $s_i$  为  $(x_i, y_i)$ , 表示点的坐标。为了消除噪声对轮廓造成的影响, 采用高斯滤波器分别对  $(x_1, x_2, \dots, x_n)$  及  $(y_1, y_2, \dots, y_n)$  进行平滑处理。

设  $\theta$  表示某一采样点的内角值, 若  $\theta > \alpha_1$ , 定义该点为凹角点; 若  $\theta < \alpha_2$ , 定义该点为凸角点; 若  $\alpha_2 \leq \theta \leq \alpha_1$ , 定义该点为平滑点。 $\alpha_1$  和  $\alpha_2$  为预先设定的阈值。实验表明,  $\alpha_1 \in [1.15\pi, 1.25\pi]$  及  $\alpha_2 \in [0.7\pi, 0.85\pi]$  时均可取得较好的效果。同时规定, 如果轮廓相邻采样点同为凸角点或凹角点, 则仅保留这些角点中内角最小或最大的一个, 其他角点作为平滑点处理。图 2.20 给出了两幅示例轮廓, 图 2.21 给出了示例轮廓角点提取结果, 其中, “ $\nabla$ ” 表示凹角点, “ $\square$ ” 表示凸角点, “ $*$ ” 表示平滑点。



图 2.20 示例轮廓

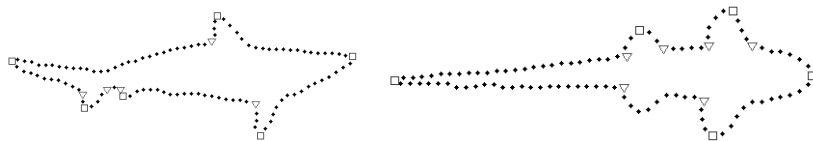


图 2.21 角点提取结果

#### (2) 基于 CSS 的角点检测

前面我们介绍了一种成功应用于形状特征描述的算子, 即曲率尺度空间描述符 (CSSD)。同时, 利用曲率尺度空间也可进行角点的检测, Mokhtarian 等提出了一种基于这种原理的检测方法<sup>[56]</sup>。这种方法的特点是, 在较大的尺度下用曲率公式计算出图像轮廓某点处的曲率, 找出局部极值点, 再通过阈值技术来检测角点, 最后在较小的尺度下对检测出的角点进行定位。

首先找到轮廓上的 T 型交叉点, 标记为 T 型角点, 然后以高斯函数的参数  $\sigma$  为尺度因子, 在一个较大的尺度下计算轮廓曲线上任意一点处的曲率, 即

$$k(u, \sigma) = \frac{X_u(u, \sigma)Y_{uu}(u, \sigma) - X_{uu}(u, \sigma)Y_u(u, \sigma)}{[X_u(u, \sigma)^2 + Y_u(u, \sigma)^2]^{3/2}} \quad (2-84)$$

其中,

$$X_u = x(u) \otimes g_u(u, \sigma), \quad X_{uu} = x(u) \otimes g_{uu}(u, \sigma) \quad (2-85)$$

$$Y_u = y(u) \otimes g_u(u, \sigma), \quad Y_{uu} = y(u) \otimes g_{uu}(u, \sigma) \quad (2-86)$$

其中,  $\otimes$  是一个卷积符号;  $u$  为弧长参数;  $g(u, \sigma)$  为高斯函数;  $g_u(u, \sigma)$  和  $g_{uu}(u, \sigma)$  分别是  $g(u, \sigma)$  关于  $u$  求一阶和二阶导数。然后把局部曲率最大点作为候选角点, 如果某个候选角点处的曲率值大于阈值  $t$ , 并且大约是相邻局部曲率最小点处曲率值的 2 倍, 则把该角点作为正确角点。最后在较小的尺度下定位这些角点, 并和 T 型角点进行比较, 剔除相隔较近的两个角点中的一个角点。

CSS 算法获得了比较好的角点检测结果, 并且对噪声也不是很敏感, 但是该算法仍有 3 方面的问题: 第一, 在检测角点的过程中使用的是单一的大尺度, 这样很容易漏掉一部分正确的角点; 第二, 当尺度过大时不能检测到真正的角点; 第三, 用于决定角点取舍的全局阈值  $t$  是预先固定的, 它的取值对最终的结果影响很大。因此, Mokhtarian 等人对轮廓曲线长度不同的图像选用不同的尺度来减少漏掉角点的情况, 然而合适的尺度并不能仅仅由图像的轮廓曲线长度所决定, 并且轮廓曲线的长度也受边缘检测算法的影响。He 等人提出了 ACSS 算法<sup>[72]</sup>, 先在小的尺度下计算曲率获取初始的候选角点, 然后利用取局部阈值和角度估计的方法移除圆角点和量化噪声和边缘细节产生的错角点, 但是引入了过多的阈值参数, 使得控制的难度和计算的复杂度都比较大。图 2.22 给出了采用 CSS 和 ACSS 检测方法提取示例轮廓角点的结果, 其中“□”表示检测出的角点。

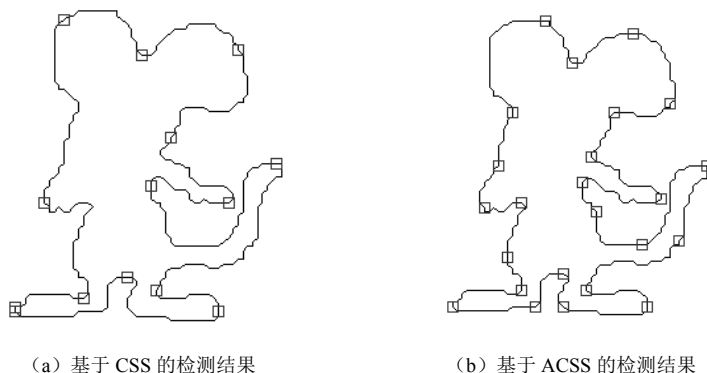


图 2.22 两种角点检测结果

在基于尺度空间的特征检测中, 融合多尺度信息的方法有多种: 一种是大尺度检测小尺度跟踪定位, 如上所述的 CSS 方法; 另一种就是, 利用多尺度乘积来增强特征信息, 同时抑制噪声对特征的影响。张小洪等人则根据多尺度积的思想, 提出了一种增强角点曲率的方法 MSCP<sup>[73]</sup>, 根据式 (2-84) 就可以计算出在第  $j$  尺度下的曲率  $k(u, \sigma_j)$ , 其多尺度曲率积为

$$P_N(u) = \prod_{j=1}^N k(u, \sigma_j) \quad (2-87)$$

利用这种方法, 取曲率的乘积大于某个阈值  $k$  的局部极大值点作为角点。图 2.23 给出了采用 MSCP 思想对图 2.22 示例轮廓在不同尺度空间下的曲率乘积的曲线图。根

据  $N$  取 1、2、3、4、5 时曲率乘积的分布，我们取适当阈值  $|k| < 0.0004$ ，可检测出示例轮廓的角点，如图 2.24 所示。

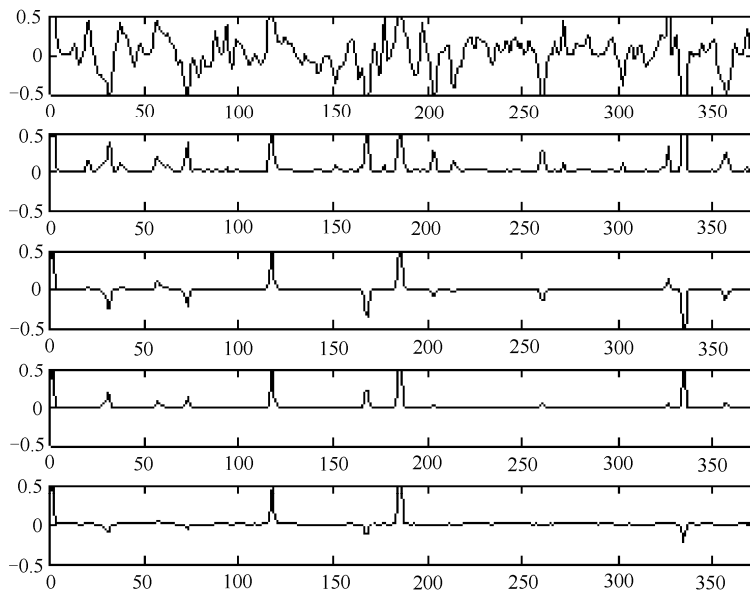


图 2.23 当  $N$  分别为 1、2、3、4、5 时的曲率乘积

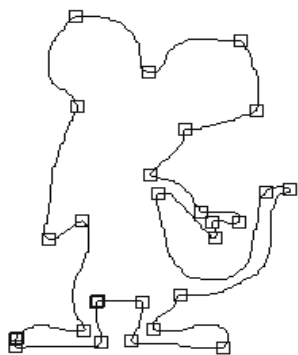


图 2.24 基于 MSCP 的检测结果

### (3) 基于 MCP 的角点检测

在充分利用多尺度信息和角点的尺度不确定性的基础上，我们给出了一种基于多尺度的曲率多项式（Multi-scale Curvature Polynomial, MCP）检测方法<sup>[74]</sup>。此方法在增强角点极大值、抑制噪声和冗余的同时，也增强了部分正确角点对应的曲率局部极大值，同时考虑了角点不同的特性，将检测到的角点区别为凸角点和凹角点。

设  $g(u, \sigma_j)$  是高斯函数  $g(u)$  在不同尺度  $\sigma_j$  下的形式，可表示为

$$g(u, \sigma_j) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_j^2}}, \quad j=1, 2, \dots \quad (2-88)$$

然后根据前述方法, 可获得不同尺度下的轮廓和轮廓上各点对应的曲率值。由于不同的尺度对曲率的作用程度不同, 对不同角点的作用程度也不同, 因此, 我们分别对不同尺度下局部极大值点对应的曲率采用加权和, 而对非极值点采用曲率积的形式, 达到增强角点的同时平滑噪声和冗余细节。可将第  $j$  尺度下轮廓上某点对应的曲率  $E_N^{(j)}(u)$  表示为

$$E_N^{(j)}(u) = \begin{cases} E_N^{(j-1)}(u) + k(u, \sigma_j), & \text{第 } j \text{ 尺度下 } u \text{ 处为局部极值} \\ E_N^{(j-1)}(u) * k(u, \sigma_j), & \text{第 } j \text{ 尺度下 } u \text{ 处不是局部极值} \end{cases} \quad (2-89)$$

其中,  $j=2, 3, \dots, N$ ;  $E_N^{(1)}(u) = k(u, \sigma_1)$ 。我们可以在连续的一些尺度空间上根据此式计算出每点的曲率多项式特征。在计算出轮廓上所有点的曲率多项式值后, 我们能够通过取阈值的方法提取出角点。由式 (2-89) 可知, 我们计算出的曲率多项式值是正负可分的, 分别代表着轮廓的凹凸性, 因此可根据此判断提取出的角点的凹凸性。又因为不同图像的轮廓体现的凹凸性不同, 在取阈值提取角点时, 应对正的局部极大值和负的局部极小值分别对待。根据实验获知, 正阈值  $E_+$  取值范围为 (0.1, 0.5), 负阈值  $E_-$  取值范围为 (-1, -0.1)。对图 2.22 示例轮廓进行检测, 图 2.25 列出了当  $N$  分别为 1、2、3、4、5 时的多项式曲率分布。图 2.26 给出了当正阈值  $E_+$  取 0.12、负阈值  $E_-$  取 -0.12 时的检测结果, 其中, “▽” 表示凹角点, “□” 表示凸角点。

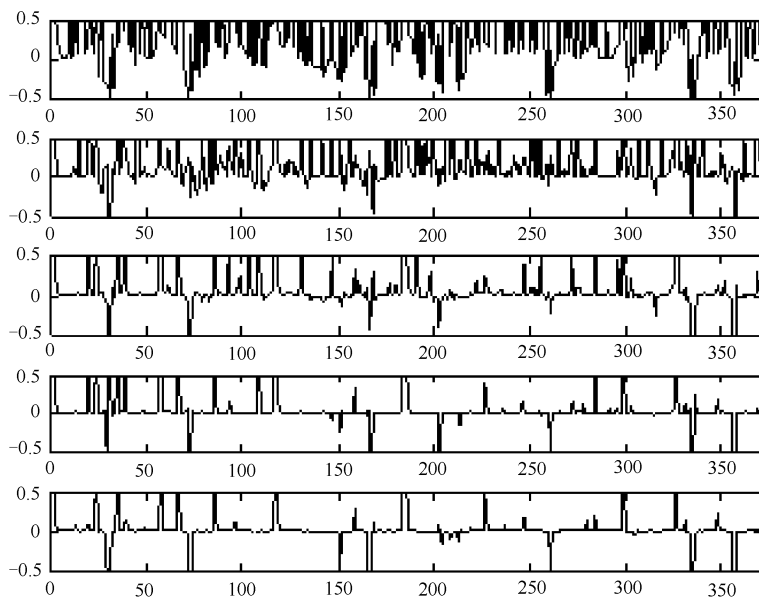


图 2.25 当  $N$  分别为 1、2、3、4、5 时的多项式曲率分布

图 2.27 给出了将 5 种检测方法应用于 airplane 图像的角点检测结果, 其中, “▽”表示凹角点, “□”表示凸角点。

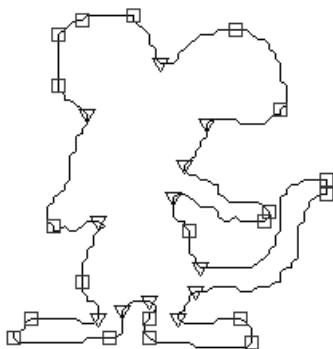


图 2.26 基于 MCP 的检测结果

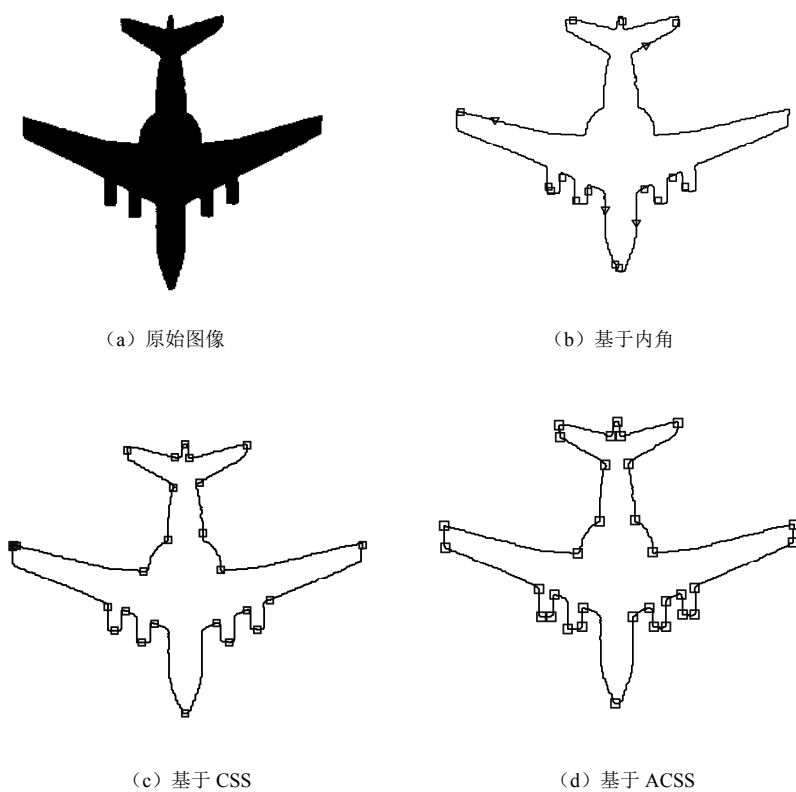


图 2.27 airplane 图像角点检测结果

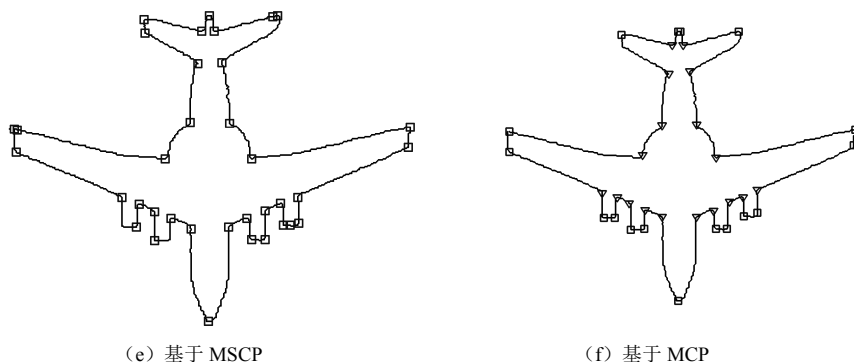


图 2.27 airplane 图像角点检测结果 (续)

## 2) 角点特征描述

### (1) 距离直方图

距离直方图定义为计算轮廓点同其质心间的距离并将其量化到一定的区间，设量化等级为  $L$ ，量化到每一区间的轮廓点数为  $d_i (i = 0, 1, \dots, L-1)$ ，则距离直方图表示为

$$\mathbf{D} = (d_0, d_1, \dots, d_{L-1}) \quad (2-90)$$

显然，上述距离直方图具有平移及旋转不变性，但不具备尺度不变性。在上述距离直方图的基础上，针对轮廓角点及平滑点，我们分别给出了其距离直方图。同时，为了获取尺度不变性，我们对其进行归一化处理，设归一化后的距离直方图表示为

$$\mathbf{H}^{(k)} = (h_0^{(k)}, h_1^{(k)}, \dots, h_{L-1}^{(k)}) \quad (2-91)$$

其中， $h_i^{(k)} = c_i^{(k)} / n_k$ ， $n_k (k = 1, 2, 3)$  分别表示凸角点、凹角点及平滑点的数目， $c_i^{(k)} (k = 1, 2, 3)$  分别表示量化到区间  $i$  的凸角点、凹角点及平滑点的数目。

然而，距离直方图仅考虑了轮廓点同其质心间的距离分布统计特性，并未考虑角点相互间的分布特性及平滑点的分布特性。为了解决该问题，有文献提出采用相对位置分布及相关单元熵来进一步描述角点及平滑点的特征。

### (2) 相对位置分布

从图 2.28 中可以看出，同类角点（凸角点或凹角点）间的距离关系也是反映角点分布的一个重要特征，这里我们就采用角点间的这种距离关系来进一步描述角点特征。

设  $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_m\}$  表示某轮廓所有凸角点的集合， $\mathbf{a}_i$  为  $(x_i, y_i)$ ，表示点的坐标， $m$  表示凸角点的数目。设  $d_{ij}$  表示相邻两个凸角点间的欧氏距离，其中， $i = 1, 2, \dots, m$ ， $j = (i+1) \text{ MOD } m$ 。这里我们取距离均方差  $\delta$  作为相对位置分布特征。同理，我们也可以得到凹角点的相对位置分布特征。



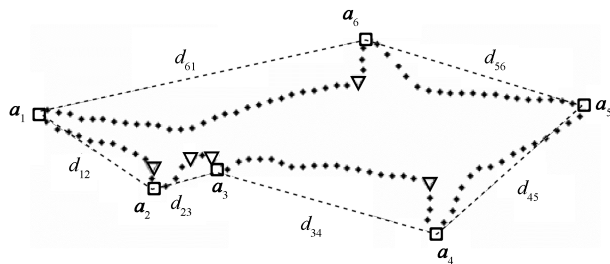


图 2.28 相对位置分布

显然, 相对位置分布特征 $\delta$ 具有旋转及平移不变性, 为了获取尺度不变性, 可对其进行归一化处理, 从而将相对位置分布定义为

$$\text{RAD} = \delta / d_{\max} \quad (2-92)$$

其中,  $d_{\max} = \max d_{ij}$ 。

### (3) 相关单元熵

从图 2.28 中也可看出, 轮廓凹凸点的变化反映了轮廓的主要特征, 但平滑点也是反映轮廓特征的重要因素。这里又提出采样相关单元熵进一步描述平滑点特征。

设 $M$ 表示轮廓角点的数目(包括凸角点和凹角点), 这里将相邻两个角点间的平滑点称为一个单元, 单元内平滑点的数目用 $l_i$ 表示,  $i=1, 2, \dots, M$ , 轮廓所有平滑点的数目用 $K$ 表示。结合空间分布熵的定义, 我们给出了相关单元熵的定义, 即

$$\text{RUE} = -\sum_{i=1}^M p_i \log_2(p_i) \quad (2-93)$$

其中,  $p_i = l_i / K$ ,  $i=1, 2, \dots, M$ 。

## 7. 基于矩的描述方法

目前, 矩技术已经广泛地应用到图像处理、计算机视觉及模式识别技术中, 如景物匹配、图像重建、图像压缩、对称性检测、图像规格化、纹理分割、边缘检测、目标识别和图像检索等。并且许多矩的快速算法也被提了出来。下面主要讨论基于边缘的矩描述方法。

### 1) 轮廓矩

物体的轮廓是描绘物体形状的重要特征, 可以将其看成是一条具有方向性的相互衔接的链条。在利用不变矩对其进行描述时, 应首先研究各个不变矩的含义和物理体现, 并结合边界轮廓的特性, 使其更充分地描述轮廓特征。

利用几何矩进行基于区域的特征描述时, 不同的阶数表示了不同的特性。针对轮廓矩来说, 各阶几何矩的物理意义可以表述如下。

### (1) 零阶几何矩

根据几何矩的定义，图像轮廓的零阶几何矩  $M_{00}$  定义为

$$M_{00} = \iint_C dx dy \quad (2-94)$$

由上式可见，对于边界轮廓，几何矩  $M_{00}$  表示边界点数目的总和，即周长。

### (2) 一阶几何矩

一阶几何矩  $M_{01}$  及  $M_{10}$  分别是图像关于  $x$  轴和  $y$  轴的矩，可以用来确定边界轮廓的几何中心点，其坐标  $(\bar{x}_0, \bar{y}_0)$  可以表示为

$$\bar{x}_0 = \frac{M_{10}}{M_{00}}, \quad \bar{y}_0 = \frac{M_{01}}{M_{00}} \quad (2-95)$$

通常，将坐标系原点移至轮廓形状的几何中心点的矩，称为几何中心矩。这一变化使矩的计算独立于图像的坐标系。其中，几何中心矩  $\mu_{pq}$  可表示为

$$\mu_{pq} = \iint (x - \bar{x}_0)^p (y - \bar{y}_0)^q dx dy \quad (2-96)$$

### (3) 二阶几何中心矩

由式 (2-96) 可以得出轮廓的二阶矩和三阶矩，其中二阶矩包括  $\mu_{11}$ 、 $\mu_{20}$  和  $\mu_{02}$ 。 $\mu_{11}$  表示轮廓的倾斜度， $\mu_{11}$  大于零表示图像向左上倾斜，小于零表示向右上倾斜。 $\mu_{20}$  和  $\mu_{02}$  分别表示轮廓在水平和垂直方向上的伸展均衡度， $\mu_{20}$  大于零表示图像下部的水平伸展比图像上部大，小于零表示上部的水平伸展比下部大； $\mu_{02}$  大于零表示图像右边的垂直伸展比左边大，小于零表示左边的垂直伸展比右边大。另外， $\mu_{20}$  和  $\mu_{02}$  为对  $x$  轴和  $y$  轴的惯性矩，通过组合它们可以确定几个重要的特性，如主轴比和方向、椭圆性等。

### (4) 三阶矩或三阶以上矩

三阶矩或三阶以上矩，则是轮廓细节和低阶矩变化的具体表现。其中， $\mu_{30}$  和  $\mu_{03}$  描述了边界曲线投影的扭曲程度，即关于均值对称分布的偏差程度，分别表示轮廓在水平和垂直方向上的重心偏移度。 $\mu_{30}$  大于零表示重心偏左，小于零表示重心偏右； $\mu_{03}$  大于零表示重心向上偏移，小于零表示重心向下偏移。 $\mu_{21}$  和  $\mu_{12}$  表示轮廓的水平与垂直伸展的均衡程度， $\mu_{21}$  大于零表示轮廓上部的水平伸展比下部大，小于零表示下部的水平伸展比上部大； $\mu_{12}$  大于零表示轮廓右边的垂直伸展比左边大，小于零表示轮廓左边的垂直伸展比右边大。

## 2) Chen 不变矩

自从 Hu 提出不变矩以后，不变矩在形状识别和分类中获得了广泛的应用。但不变矩是针对区域像素计算的，它不能用于边界的检测。为使其可用于边界形状的检测，Chen<sup>[75]</sup>在区域不变矩的基础上，提出了矩的轮廓描述，其形式表示为

$$M_{pq} = \int_C x^p y^q ds \quad (2-97)$$

其中， $p, q = 0, 1, 2, 3, \dots$ ； $\int_C$  表示在曲线  $C$  上积分； $ds = \sqrt{(dx)^2 + (dy)^2}$ 。对于数字轮廓，

轮廓矩可表示为

$$M_{pq} = \sum_{(x,y) \in C} x^p y^q \quad (2-98)$$

轮廓中心矩定义为

$$\mu_{pq} = \sum_{(x,y) \in C} (x - \bar{x})^p (y - \bar{y})^q \quad (2-99)$$

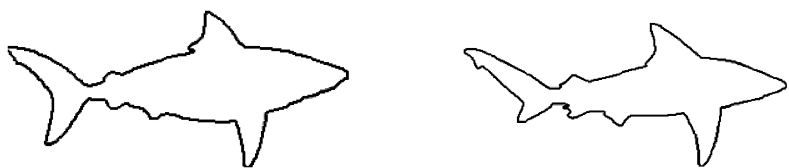
其中,  $\bar{x} = M_{10}/M_{00}$ ;  $\bar{y} = M_{01}/M_{00}$ 。上述不变矩具有旋转不变性, 为了获取尺度不变性, 归一化的中心矩可以表示为

$$\eta_{pq} = \mu_{pq} / \mu_{00}^\gamma \quad (2-100)$$

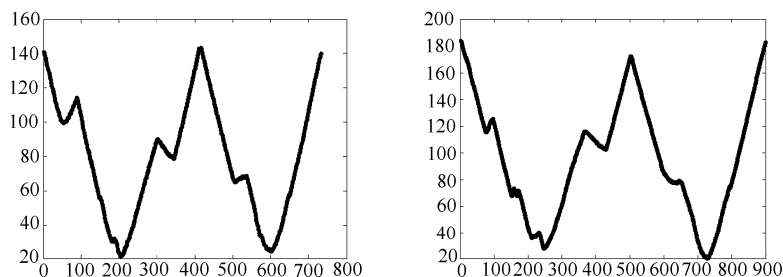
其中,  $\gamma = p + q + 1$ 。轮廓矩的 7 个不变量的计算类似于 Hu 不变矩的 7 个不变量, 这里不再详细给出。

### 3) 边界序列矩

一般物体的轮廓能够通过分段的多项式拟合、链码、梯度角描述, 那么轮廓就可以表示为一个有序的几何向量。此描述方法先假设边界点为  $\{(x(i), y(i)) | i = 1, 2, \dots, N\}$ , 边界点的数目为  $N$ , 由边界点围成的中心点即为  $(\bar{x}, \bar{y})$ , 则可将每点与中心点的距离  $z$  作为此边界的一维描述函数  $z(i) (i = 1, 2, \dots, N)$ 。图 2.29 (a) 给出了两幅鱼类示例轮廓, 图 2.29 (b) 给出了它们对应的  $z(i)$  函数。



(a) 示例轮廓



(b) 对应的  $z(i)$  函数曲线图

图 2.29 一维描述函数示例

利用此一维描述函数, 可以获取边界轮廓的几何矩<sup>[76]</sup>, 即

$$M_p = \frac{1}{N} \sum_{i=1}^N [z(i)]^p, \quad p=1,2,\dots \quad (2-101)$$

$p$  阶中心矩为

$$\mu_p = \frac{1}{N} \sum_{i=1}^N [z(i) - M_1]^p \quad (2-102)$$

选取 4 个低阶矩的组合, 可用于目标的识别与分类, 且这 4 个低阶矩具有尺度、旋转和平移不变性, 其形式可表示为

$$\left. \begin{aligned} S_1 &= \mu_2^{1/2} / M_1 \\ S_2 &= \mu_3 / \mu_2^{3/2} \\ S_3 &= \mu_4 / \mu_2^2 \\ S_4 &= \mu_5 / \mu_2^{5/2} \end{aligned} \right\} \quad (2-103)$$

其中,  $S_1$  称为归一化的幅度变化量, 它是一个非负数, 当且仅当边界为圆时  $S_1 = 0$ ;  $S_2$  为与形状的对称程度有关的量, 称为歪斜度;  $S_3$  为与密度函数的峰值峭度有关的量, 称为峭度。实验证明, 利用这 4 个参数可实现对形状的可靠识别和分类。同时在边界序列矩的基础上, 我们导出了 3 个具有尺度、旋转及平移不变性的矩, 表示为<sup>[74]</sup>

$$\left. \begin{aligned} T_1 &= \mu_2 / M_1^2 \\ T_2 &= \mu_3 / M_1^3 \\ T_3 &= \mu_4 / M_1^4 \end{aligned} \right\} \quad (2-104)$$

#### 4) 极半径不变矩

极半径矩由曹茂永等<sup>[77]</sup>提出, 这里我们主要介绍其对轮廓曲线的描述矩。设  $p(x,y)$  为边界上的任一点,  $r$  为该点到形心  $(\bar{x}, \bar{y})$  的距离, 那么第  $p$  阶极半径矩和中心矩的离散形式定义为

$$\left. \begin{aligned} m_{np} &= \frac{1}{N} \int \left( \frac{r}{\bar{r}} \right)^p ds \\ m_{ncp} &= \frac{1}{N} \int \left( \frac{r - \bar{r}}{\bar{r}} \right)^p ds \end{aligned} \right\} \quad (2-105)$$

其中,  $N$  为边界的周长;  $\bar{r} = \frac{1}{N} \int r ds$  为平均极半径;  $ds = r d\theta$  为极坐标下的线积分元;

当边界的形状均匀缩放  $\alpha$  倍时, 周长变为  $\alpha N$ 。可见, 此极半径矩和中心矩具有尺度、旋转和平移不变性。 $m_{np}$  和  $m_{ncp}$  为不变量, 那么它们的组合也必是不变量, 选取适当的组合可用于物体的识别与匹配。从中选取 5 个低阶矩的组合, 可用于目标的识别与分类, 且这 5 个低阶矩具有尺度、旋转和平移不变性。对于离散的数字轮廓, 用求和代替积分, 设取轮廓上所有的  $N$  个点为采样点, 则 5 个不变矩可表示为

$$V_1 = m_{n2} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i}{\bar{r}} \right)^2 \quad (2-106)$$

$$V_2 = m_{nc2} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i - \bar{r}}{\bar{r}} \right)^2 \quad (2-107)$$

$$V_3 = m_{nc3} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i - \bar{r}}{\bar{r}} \right)^3 \quad (2-108)$$

$$V_4 = m_{nc4} = \frac{1}{N} \sum_{i=1}^N \left( \frac{r_i - \bar{r}}{\bar{r}} \right)^4 \quad (2-109)$$

$$V_5 = \frac{m_{nc2}}{\sqrt{m_{nc4}}} = \frac{1}{\sqrt{N}} \frac{\sum_{i=1}^N (r_i - \bar{r})^2}{\sqrt{\sum_{i=1}^N (r_i - \bar{r})^4}} \quad (2-110)$$

### 2.2.3 基于区域的描述方法

基于区域的描述方法将区域形状当作一个整体来看待,该方法有效地利用了区域内的所有像素,因而受噪声和形状变化的影响相对较小。同基于轮廓的形状描述方法一样,基于区域的形状描述方法也分为全局型和局部型两种类型。形状的区域特征主要有区域的面积、欧拉数、离散度、偏心率、区域不变矩、区域骨架、几何不变矩、Legendre 矩、Zernike 矩、伪 Zernike 矩、旋转矩、复数矩、通用傅里叶描述符、角半径变换等方法。

本小节主要介绍简单几何参数描述符、几何不变矩、Zernike 矩、角半径变换、通用傅里叶描述符、基于状态矩阵的描述方法、基于平坦度及凹凸度的描述方法、基于信息熵的描述方法。

#### 1. 简单几何参数描述符

##### 1) 区域面积

区域面积是区域的一个基本特征,它描述区域的大小。设正方形像素的边长为单位长,则计算区域面积就是对属于区域的像素计数。

##### 2) 区域重心

区域重心是一种全局描述符,区域重心的坐标是根据所有属于区域的点计算出来的,其计算公式为

$$\left. \begin{aligned} \bar{x} &= \frac{1}{A} \sum_{(x,y) \in R} x \\ \bar{y} &= \frac{1}{A} \sum_{(x,y) \in R} y \end{aligned} \right\} \quad (2-111)$$

其中,  $R$  表示某一区域;  $A$  表示区域面积 (即区域像素数)。

### 3) 区域灰度 (密度)

目标的灰度特性要结合原始灰度图和分割图来得到。常用的区域灰度特征有目标灰度的最大值、最小值、中值、平均值、方差及高阶矩等统计量, 它们多可借助灰度直方图得到。

### 4) 欧拉数

欧拉数是一种拓扑描述符, 区域的拓扑性质对区域的全局描述很有用, 这种性质既不依赖距离, 也不依赖基于距离测量的其他特性。设  $H$  表示区域内的空数,  $C$  表示区域内的连通组元的个数, 则欧拉数可定义为

$$E = C - H \quad (2-112)$$

### 5) 形状参数

形状参数是根据区域的周长和区域的面积计算出来的, 其计算公式为

$$F = \frac{\|B\|^2}{4\pi A} \quad (2-113)$$

其中,  $B$  为区域的周长。可以看出, 圆形的形状参数为 1, 当区域为其他形状时, 形状参数大于 1。形状参数在一定程度上描述了区域的紧凑性 (compactness), 且对尺度变换及旋转变换不敏感。

### 6) 偏心率

偏心率 (eccentricity) 也可叫伸长度 (elongation), 它也在一定程度上描述了区域的紧凑性。由惯性推出的偏心率计算公式为

$$E = \sqrt{\frac{(A+B) - \sqrt{(A-B)^2 + 4H^2}}{(A+B) + \sqrt{(A-B)^2 + 4H^2}}} \quad (2-114)$$

其中,  $A = \sum m_i (y_i^2 + z_i^2)$ 、 $B = \sum m_i (x_i^2 + z_i^2)$ 、 $C = \sum m_i (x_i^2 + y_i^2)$  分别是刚体绕  $x$ 、 $y$ 、 $z$  轴的转动惯量; 而  $F = \sum m_i y_i z_i$ 、 $G = \sum m_i z_i x_i$ 、 $H = \sum m_i x_i y_i$  均称作惯性积。

### 7) 球状性

二维区域的球状性 (sphericity) 的计算需要用到区域重心, 它定义为

$$S = \frac{r_i}{r_c} \quad (2-115)$$

其中,  $r_i$  代表区域内切圆的半径; 而  $r_c$  代表区域外接圆的半径; 两个圆的圆心都在区域重心上。球状性的值当区域为圆时达到最大值 1, 而当区域为其他形状时则小于 1。它也不受平移、旋转和尺度变化的影响。

## 8) 圆形性

圆形性 (circularity) 是用区域的所有边界点定义的特征量, 其计算公式为

$$C = \frac{\mu_R}{\sigma_R} \quad (2-116)$$

其中,  $\mu_R = \frac{1}{K} \sum_{k=0}^{K-1} \|(x_k, y_k) - (\bar{x}, \bar{y})\|$  为从区域重心到边界点的平均距离;

$\sigma_R = \frac{1}{K} \sum_{k=0}^{K-1} [\|(x_k, y_k) - (\bar{x}, \bar{y})\| - \mu_R]^2$  为从区域重心到边界点的距离的均方差;  $K$  为边界像素点的个数。当区域趋向圆形时, 圆形性是单增趋向无穷的, 同时圆形性也不受区域的平移、旋转和尺度变化的影响。

## 2. 几何不变矩

图像的矩函数在模式识别、目标分类中得到了广泛的应用。1962 年, Hu<sup>[78]</sup>提出了图像识别的不变矩理论, 首次提出了基于代数不变量的矩不变量, 并通过对几何矩的非线性组合, 导出了一组对于图像平移、旋转和尺度变化不变的矩。不变矩是图像的一种统计特征, 它利用图像灰度分布的各阶矩来描述图像灰度的分布特性。

离散数字图像  $f(x, y)$  的  $p+q$  阶矩定义为

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (2-117)$$

其  $p+q$  阶中心矩定义为

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (2-118)$$

其中,  $\bar{x} = m_{10}/m_{00}$ 、 $\bar{y} = m_{01}/m_{00}$  表示图像的区域重心。中心矩表示了图像内不同灰度级的像素相对于其重心是如何分布的, 因此中心矩具有位置无关性。为了获取针对图像缩放无关的性质, 可以对该中心矩进行规格化操作, 规格化后的中心矩表示为

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad (2-119)$$

其中,  $\gamma = \frac{p+q}{2} + 1$ ,  $p+q = 2, 3, \dots$ 。中心规格矩对于图像缩放、平移和旋转均保持不变 (夏德深等 1997)。基于规格化的二阶和三阶中心矩, 可以导出下面 7 个矩组。

$$\begin{aligned}
 \phi_1 &= \eta_{20} + \eta_{02} \\
 \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
 \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
 \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
 \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
 &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
 \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
 \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\
 &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]
 \end{aligned}$$

上述 7 个不变矩组称为 Hu 不变矩。Hu M.K. 证明了这个矩组中的  $\phi_1 \sim \phi_6$  具有尺度、平移和旋转不变性，而  $\phi_7$  只具有尺度和平移不变性，不具有旋转不变性，仅在镜面对称时保持不变。在 Hu 不变矩的基础上，文献[79]、[80]、[81]等对不变矩进行了推广。

### 3. Zernike 矩

Zernike 矩定义为<sup>[82]</sup>

$$Z_{nm} = \frac{n+1}{\pi} \sum_y \sum_x V_{nm}^* f(x, y), \quad x^2 + y^2 \leq 1 \quad (2-120)$$

其中， $V_{nm}(x, y) = V_{nm}(\rho \cos \theta, \rho \sin \theta) = R_{nm}(\rho) \exp(jm\theta)$ 。  $R_{nm}(\rho)$  的定义为

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! [(n+|m|)/2-s]! [(n-|m|)/2-s]!} \times \rho^{n-2s} \quad (2-121)$$

其中， $n$  和  $m$  为非负整数，并且必须满足  $n - |m|$  为偶数且  $n \geq |m|$ 。

自从 Zernike 矩的概念被引入以来，Zernike 矩便因其具有优良的旋转不变性而在模式识别等领域得到广泛的应用<sup>[83,84]</sup>。在图像分析中，由于 Zernike 多项式的正交性，可以使信息冗余达到最优，它的递归性质使得矩的快速算法成为可能。Zernike 矩还有一个重要的特性：图像旋转一定角度后的 Zernike 矩与原图像的 Zernike 矩有非常简单的关系，就是图像旋转后的 Zernike 矩仅仅相位发生变化而幅值保持不变。

除了上述的几何不变矩及 Zernike 矩之外，还有其他很多区域矩的描述方法，有关矩的描述基本上分为两大类型：正交矩及非正交矩。常用的正交矩有 Legendre moments、Zernike moments、pseudo-Zernike moments 等，常用的非正交矩有 geometric moments（即几何不变矩）、complex moments、rotation moments 等。

### 4. 角半径变换

角半径变换（Angular Radial Transformation, ART）是 MPEG-7 推荐的另一个基于区域的形状描述符，也是一种基于矩的图像描述符。它使用一组 ART 系数，描述



单个连通区域或者多个不连通区域，并且对旋转具有鲁棒性。ART 是定义在极坐标下的一个单位圆内的二维复变换，是一种正交变换，对噪声具有鲁棒性<sup>[85]</sup>。ART 系数定义为

$$F_{nm} = \langle V_{nm}(\rho, \theta) f(\rho, \theta) \rangle = \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta) \rho d\rho d\theta \quad (2-122)$$

其中， $F_{nm}$  是序数为  $(n, m)$  的 ART 系数； $f(\rho, \theta)$  是在极坐标上的一个图像灰度方程； $V_{nm}(\rho, \theta)$  是 ART 的基本函数，如图 2.30 所示； $V_{nm}^*(\rho, \theta)$  为其复共轭。ART 的基本函数沿着角方向和放射方向是可分离的，它的表达方式为

$$V_{nm}(\rho, \theta) = A_m(\theta) R_n(\rho) \quad (2-123)$$

$$\text{其中, } A_m(\theta) = \frac{1}{2\pi} \exp(jm\theta); \quad R_n(\rho) = \begin{cases} 1, & n = 0 \\ 2\cos(\pi n \rho), & n \neq 0 \end{cases}.$$

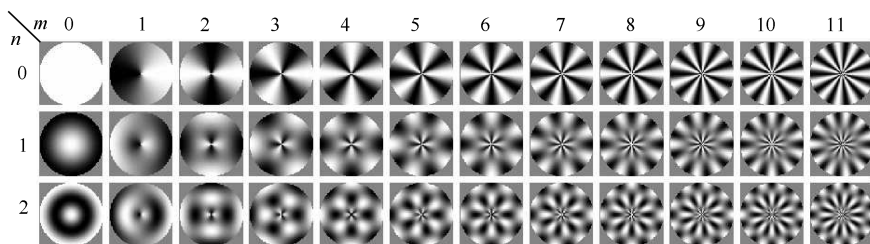
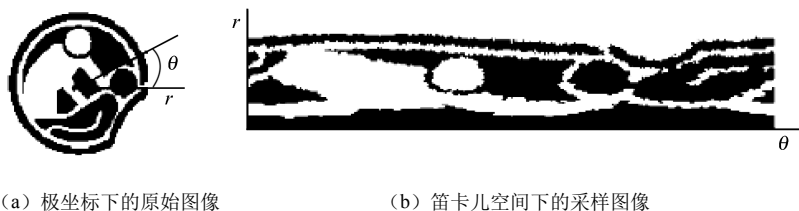


图 2.30 ART 的基本函数

## 5. 通用傅里叶描述符

通用傅里叶描述符 (Generic Fourier Descriptor, GFD)<sup>[86]</sup> 采用了修正的平面极坐标傅里叶变换，对图像进行采样，将采样的信息重新绘制在笛卡儿直角坐标系下，再对该直角坐标下的图像进行傅里叶变换，如图 2.31 所示。



(a) 极坐标下的原始图像

(b) 笛卡儿空间下的采样图像

图 2.31 通用傅里叶描述符示意图

修正的极坐标傅里叶变换定义为

$$GF(\rho, \theta) = \sum_r \sum_i f(r, \theta_i) \exp \left[ j2\pi \left( \frac{r}{R} \rho + \frac{2\pi i}{T} \phi \right) \right] \quad (2-124)$$

其中,  $0 \leq r = \left[ (x - x_c)^2 + (y - y_c)^2 \right]^{\frac{1}{2}} \leq R$ ;  $\theta_i = i(2\pi/T)$  ( $0 \leq i \leq T$ );  $(x_c, y_c)$  是形状的中心点;  $R$ 、 $T$  分别为半径和角度的分辨率。GFD 具有平移不变性, 为了获取尺度和旋转不变性, 可进行如下规范化。

$$\text{GFD} = \left\{ \frac{\text{GF}(0,0)}{\text{area}}, \frac{\text{GF}(0,1)}{\text{GF}(0,0)}, \dots, \frac{\text{GF}(0,n)}{\text{GF}(0,0)}, \dots, \frac{\text{GF}(m,0)}{\text{GF}(0,0)}, \dots, \frac{\text{GF}(m,n)}{\text{GF}(0,0)} \right\} \quad (2-125)$$

其中,  $\text{area}$  表示极坐标采样时形状所占的最大面积;  $m$  是最大半径频数;  $n$  是最大半径角度频数。

## 6. 基于状态矩阵的描述方法

### 1) 状态矩阵的定义

为了描述图像的形状信息, 可将图像的色调、饱和度和亮度 3 个灰度分量分别划分为  $n \times n$  个小区块子块。对于某一区域子块, 如果该区域内像素间的灰度值差别较小, 那么在人眼看来, 该区域表现得就比较平坦; 如果该区域内像素间的灰度值差别较大, 则该区域表现得就比较起伏。依据这种特性, 我们可以将图像的区域子块划分为不同的类型<sup>[87]</sup>。

设  $\sigma$  表示子块内像素灰度值的标准差,  $\alpha_1$ 、 $\alpha_2$  ( $0 < \alpha_1 < \alpha_2$ ) 为事先给定的阈值, 则有如下图像分块状态的定义。

**平坦态:** 对于图像中的某一分块, 如果满足条件  $\sigma < \alpha_1$ , 则定义该子块的状态为平坦态, 并用 0 表示该种状态。

**边缘态:** 对于图像中的某一分块, 如果满足条件  $\sigma > \alpha_2$ , 则定义该子块的状态为边缘态, 并用 2 表示该种状态。

**纹理态:** 对于图像中的某一分块, 如果满足条件  $\alpha_1 < \sigma < \alpha_2$ , 则定义该子块的状态为纹理态, 并用 1 表示该种状态。

根据上述对子块类型的定义, 可以将图像中的每一区域子块分别采用其状态值来表示, 这种采用分块的状态来表示的图像矩阵称为图像的状态矩阵。这样, 与彩色图像对应的每个灰度图像均可转化为由 3 种状态 {0,1,2} 构成的二维状态矩阵。图像的状态矩阵反映了图像形状的变化, 并且将图像的形状变化映射到图像分块状态的交替上, 这大大减少了进行形状特征提取的计算量。如何由图像的状态矩阵提取图像的形状变化信息是关键问题, 为此我们提出采用基于马尔可夫链和状态相关图两种方法来分别从图像的状态矩阵中提取图像的形状变化信息。

### 2) 基于马尔可夫链的形状特征提取

由文献[88]的分析可知, 一个马尔可夫链运动规律的概率特性取决于它的转移概率矩阵特性, 因此, 在对图像的状态矩阵进行处理时, 即利用了马尔可夫链的这种特性。

众所周知,在二维图像空间中,像素的空间位置越近,则其相关性就越强,因而对于图像的区域子块,空间位置越近,相应的相关性也越强。同样,对于图像的状态矩阵,也具有相同的特性,即空间位置越近的状态之间的相关性越强。为了从图像的状态矩阵中提取图像的形状信息,我们采用Z字形扫描的方法将图像的状态矩阵转化为一维随机状态序列 $\{y_1, y_2, y_e, \dots, y_{n \times n}\}$  ( $y_e$ 表示子块 $e$ 的状态,  $e=1, 2, 3, \dots, n \times n$ ), 该序列对应的状态空间为 $U = \{0, 1, 2\}$ 。采用Z字形扫描方法可以尽可能将图像状态矩阵中相关性强的状态联系在一起,从而更有利于提取图像的形状特征。在图像的一维状态序列中,位置越靠近的两个状态的相关性越强,若某一状态为平坦态,则其相邻的两个状态在很大程度上可能为平坦态,随着不同状态在序列中距离的增大,它们间的相关性会逐步减弱。

针对图像一维状态序列,我们作如下假设:在图像的状态序列中,任何一种状态的出现仅与该状态的前一状态有关,而与该状态更前的状态及后继状态无关,即当前子块状态是平坦态、纹理态还是边缘态仅仅与它前一子块的状态有关。在此条件下,图像状态序列满足以下几个条件。

(1) 序列的状态数目有限(状态数为3)。

(2) 对于某一幅图像,其对应的一维状态序列可唯一确定,且该状态序列同时间的变化没有关系。

(3) 进入某状态的概率仅与该状态之前的状态有关。

根据前述的马氏链的特性即C-K方程可知,对于马氏链 $\mathbf{X}$ ,其 $k$ 步转移概率由一步转移概率完全确定,因此这里仅考虑图像状态序列的一步转移概率矩阵的计算方法,针对多步转移概率矩阵,可通过一步转移概率矩阵求得。设 $r$ 及 $t$ 是图像随机状态序列中任意两个邻接状态,从而转移概率矩阵可表示为

$$\mathbf{P} = [p_{r,t}] = \begin{bmatrix} p_{0,0} & p_{0,1} & p_{0,2} \\ p_{1,0} & p_{1,1} & p_{1,2} \\ p_{2,0} & p_{2,1} & p_{2,2} \end{bmatrix} \quad (2-126)$$

其中,  $p_{r,t} = p\{y_e = t | y_{e-1} = r\} = \frac{N_{r,t}}{\sum_{z=0}^2 N_{r,z}}$ ,  $e=1, 2, 3, \dots, n \times n$ ,  $r, t \in U$ ,  $U = \{0, 1, 2\}$ 。

若将上述的邻接状态 $r$ 及 $t$ 称为一个状态对,记为 $(r, t)$ ,则式(2-126)中 $N_{r,t}$ 表示在图像随机状态序列中,状态对 $(r, t)$ 出现的总次数。依据上述方法,就可以分别计算出彩色图像3个灰度分量图像的转移概率矩阵。转移概率矩阵反映了图像状态矩阵的状态变化,这种状态变化又反映了图像形状的变化,因而可将所求出的3个转移概率矩阵作为图像形状特征的描述。

由图像状态序列所求取的转移概率矩阵同数学意义上由随机序列所求取的转移概率矩阵稍有不同。假设随机状态序列的状态空间 $U = \{0, 1, 2\}$ ,则在数学意义上,转

移概率矩阵  $[p_{r,t}]$  满足

$$\sum_{t=0}^2 p_{r,t} = 1, r, t \in U \quad (2-127)$$

但通过图像状态矩阵转化来的图像状态序列却并不完全满足上式, 由于具体到某幅图像不同灰度分量对应的随机状态序列, 其状态空间为图像库状态空间的子集, 因此, 如果某图像对应的随机状态序列中不存在某种状态, 那么采用式 (2-126) 来计算状态序列的转移概率矩阵时就会出现  $\sum_{t=0}^2 p_{r,t} = 0$  的情况, 即分母为零的情况。在实际处理过程中, 可首先判断  $\sum_{t=0}^2 p_{r,t}$  的取值, 如果  $\sum_{t=0}^2 p_{r,t} = 0$ , 则直接令  $p_{r,0} = p_{r,1} = p_{r,2} = 0$ , 以防止出现分母为零的情况。

### 3) 基于状态相关图的特征提取

为了有效地提取图像形状的空间分布特征, 可采用与颜色相关图<sup>[37]</sup>类似的方法, 利用状态相关图及状态自关联图来反映状态空间中不同状态的空间距离的分布特征。

## 7. 基于平坦度及凹凸度的描述方法

### 1) 平坦度及凹凸度的定义

对彩色图像来说, 图像的形状变化主要通过图像亮度分量中像素灰度的变化来反映。因此, 对于某一区域子块, 如果该区域内像素间的灰度值差别较小, 那么在人眼看来, 该区域表现得就比较平坦; 如果该区域内像素间的灰度值差别较大, 则该区域表现得就比较起伏。而区域内像素灰度值的这种变化可以通过分块区域内像素灰度值的标准差来反映, 因而对于某一子块, 如果其内部像素灰度值的标准差很小, 则表明该子块内像素的灰度值比较接近, 因此该子块对应图像中灰度变化平坦的区域; 如果其内部像素灰度值的标准差很大, 则表明该子块内像素的灰度值变化较大, 而只有包含图像边缘细节的区域子块才具有这种特性。

为了描述图像形状的这种变化特征, 引入两个新的图像分块形状描述符: 平坦度和凹凸度<sup>[89]</sup>。设  $v(i, j)$  表示亮度分量中像素  $(i, j)$  的灰度值, 则有如下定义。

**平坦度:** 图像亮度分量子块内所有像素灰度的均值定义为该子块的平坦度, 即

$$\text{flatness} = \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m v(i, j) \quad (2-128)$$

**凹凸度:** 为了定义分块的凹凸度, 首先将亮度分量中所有像素灰度的均值  $\text{mean} \left[ \text{mean} = \frac{1}{R_1 \times R_2} \sum_{i=1}^{R_1} \sum_{j=1}^{R_2} v(i, j), R_1, R_2 \text{ 表示图像的尺寸} \right]$  看作一个平面, 对于每一分块, 若  $\text{flatness} > \text{mean}$ , 则称该分块具有凸性; 若  $\text{flatness} < \text{mean}$ , 则称该分块具有

凹性。从而分块的凹凸度定义为

$$\text{roughness} = \begin{cases} \left\{ \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m [v(i, j) - \text{flatness}]^2 \right\}^{\frac{1}{2}}, & \text{flatness} > \text{mean} \\ 0, & \text{flatness} = \text{mean} \\ -\left\{ \frac{1}{l \times m} \sum_{i=1}^l \sum_{j=1}^m [v(i, j) - \text{flatness}]^2 \right\}^{\frac{1}{2}}, & \text{flatness} < \text{mean} \end{cases} \quad (2-129)$$

其中,  $l$ 、 $m$  表示分块的尺寸。

## 2) 形状特征量化

在提取了图像每一分块的平坦度和凹凸度特征后, 接下来需要按照平坦度和凹凸度两个属性对图像的分块进行分类, 因此就需要选用合适的量化策略对图像分块的平坦度和凹凸度进行量化。由平坦度和凹凸度的定义可以看出, 平坦度反映了图像分块的基本形状, 而凹凸度反映了图像分块的形状变化, 因此在对平坦度和凹凸度特征进行量化时需要采用不同的量化策略。对于平坦度, 采用粗量化即可; 而对于反映图像形状变化的凹凸度, 则需要采用细量化。

### (1) 平坦度量化

在对平坦度进行量化时, 采用均匀量化和非均匀量化两种量化方法, 具体可参见文献[89]。

### (2) 凹凸度量化

对于凹凸度, 由于其基本服从正态分布, 采用分区间-均匀量化和等概率量化两种量化方法。

#### ①分区间-均匀量化

由于凹凸度的分布基本服从正态分布, 设凹凸度的均值为 $\mu$ , 均方差为 $\sigma$ , 则其分布密度可表示为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2-130)$$

其分布函数可表示为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (2-131)$$

对于正态分布, 由于元素分布主要集中在均值 $\mu$ 的两侧, 因此这里提出采用分区间-均匀量化策略, 对于靠近均值 $\mu$ 的区间进行较细的均匀量化, 对于其他区间进行较粗的均匀量化, 具体量化过程如下。

- a. 计算整幅图像分块凹凸度的均值 $\mu$ 和均方差 $\sigma$ 。
- b. 由于图像凹凸度的分布基本服从正态分布, 因此凹凸度主要分布在 $[\mu - \sigma,$

$\mu + \sigma]$  的范围内, 在对凹凸度进行量化时, 针对该区域的值进行较细的均匀量化, 针对该区域之外的值进行较粗的均匀量化, 量化方法为

$$Q(i, j) = \begin{cases} \frac{\text{roughness}(i, j)}{\text{step1}}, & \text{roughness}(i, j) \in [\mu - \sigma, \mu + \sigma] \\ \frac{\text{roughness}(i, j)}{\text{step2}} + \frac{\mu + \sigma}{\text{step1}}, & \text{roughness}(i, j) \in (\mu + \sigma, 128] \\ \frac{\text{roughness}(i, j)}{\text{step2}} + \frac{\mu - \sigma}{\text{step1}}, & \text{roughness}(i, j) \in [-128, \mu - \sigma) \end{cases} \quad (2-132)$$

其中,  $Q(i, j)$  表示对分块  $(i, j)$  凹凸度的量化结果[实验中对  $Q(i, j)$  采用取整表示];  $\text{step1}$ 、 $\text{step2}$  表示量化步长。通过调整  $\text{step1}$  及  $\text{step2}$  的取值, 可以控制将分块划分为不同类型的精确度。为了详细地描述图像的形状变化,  $\text{step1}$  及  $\text{step2}$  应取相对较小的值, 若想粗略地描述图像形状的变化,  $\text{step1}$  及  $\text{step2}$  取相对较大的值即可, 同时  $\text{step1}$  及  $\text{step2}$  的取值也将会影响所提取的图像形状特征的维数变化。在量化完毕后, 根据量化结果将图像的分块划分为不同的类型, 同时采用不同类型分块的统计直方图作为图像形状特征的描述。

## ②等概率量化

虽然采用分区间-均匀量化也可取得较好的效果, 但是在每一划分的区间内部仍然属于刚性量化, 为了解决区间刚性量化的问题, 又提出了等概率的量化方法。等概率量化的基本思想是首先将概率区间  $[0, 1]$  进行均匀划分, 然后根据划分后的概率按逆方向计算划分的区间边界。

设  $X$  服从正态分布, 当  $\mu = 0$ 、 $\sigma = 1$  时称  $X$  服从的分布为标准正态分布, 设标准正态分布的概率密度和分布函数分别用  $\varphi(x)$  和  $\Phi(x)$  表示, 即有

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (2-133)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (2-134)$$

这里引进标准正态分布是因为标准正态分布函数  $\Phi(x)$  的取值可以通过查找相应的标准正态分布函数表获取, 从而避免复杂的运算, 而且对于正态分布也可通过下面的定理将其转换为标准正态分布。

因此, 在对图像的凹凸度进行量化时, 首先将其概率区间  $[0, 1]$  均匀划分为  $n_t$  个区间, 然后根据其概率分布求每一分界点所对应的值。具体过程如下。

a. 首先将凹凸度的分布由正态分布转换为标准正态分布。

b. 由于正态分布关于  $x = \mu$  对称, 因此, 只需将概率区间  $[0.5, 1]$  划分为  $\frac{n_t}{2}$  个区间即可, 从而概率区间被划分为  $\{0.5, 0.5 + 1/n_t, \dots, 0.5 + i/n_t, \dots, 1 - 1/n_t\}$ 。

c. 根据式 (2-134) 及所划分的概率区间计算相应的分界点的值。

$$\begin{aligned}
 \Phi\left(\frac{x_1 - \mu}{\sigma}\right) &= 0.5 \Leftrightarrow x_1 = \mu + \sigma\Phi^{-1}(0.5) = \mu \\
 \Phi\left(\frac{x_2 - \mu}{\sigma}\right) &= 0.5 + \frac{1}{n_r} \Leftrightarrow x_2 = \mu + \sigma\Phi^{-1}\left(0.5 + \frac{1}{n_r}\right) \\
 &\dots\dots\dots \\
 \Phi\left(\frac{x_i - \mu}{\sigma}\right) &= 0.5 + \frac{i}{n_r} \Leftrightarrow x_i = \mu + \sigma\Phi^{-1}\left(0.5 + \frac{i}{n_r}\right) \\
 &\dots\dots\dots \\
 \Phi\left(\frac{x_{\frac{n_r}{2}} - \mu}{\sigma}\right) &= 1 - \frac{1}{n_r} \Leftrightarrow x_{\frac{n_r}{2}} = \mu + \sigma\Phi^{-1}\left(1 - \frac{1}{n_r}\right)
 \end{aligned}$$

而  $\Phi^{-1}$  的取值可以通过查找标准正态分布函数表取得, 从而降低了计算的复杂度。

d. 根据正态分布的对称特性, 以及上述计算得到的分界点, 即可得到在概率区间  $[0, 0.5]$  上的分界点。

量化完成后, 设平坦度被量化为  $n_r$  个级别, 从而图像的分块被划分为  $n_f \times n_r$  个类别。按照以上的量化级数, 我们把平坦度 (flatness) 和凹凸度 (roughness) 两个分量合并为一个一维的特征向量, 即

$$h_s = n_r \times \text{flatness}' + \text{roughness}' \quad (2-135)$$

其中,  $\text{flatness}'$  和  $\text{roughness}'$  为平坦度和凹凸度的量化结果。这样平坦度 (flatness) 和凹凸度 (roughness) 两个分量在一维向量上分布开来。根据式 (2-135),  $h_s$  的取值范围为  $[0, n_f \times n_r - 1]$ , 也就是说, 根据量化后的结果可以统计得到  $n_f \times n_r$  柄的一维直方图, 该直方图即可用来描述形状特征。

## 8. 基于信息熵的描述方法

对数字图像而言, 不同灰度的像素出现次数的不同及其空间分布位置的不同, 使得图像呈现不同的形状。因此, 不同形状的图像所包含的熵也是不尽相同的, 故而可以用熵描述图像的形状特征。这里称为 EBIR (Entropy-Based Image Retrieval) 算法。

### 1) 图像信息熵的定义

按照 2.1.3 小节中的图像颜色的信息熵的定义, 设  $M$  表示图像灰度级集合, 向量  $(p_1, p_2, \dots, p_m, \dots)$  为图像灰度直方图, 其信息熵定义为

$$e = - \sum_{m \in M} p_m \log_2(p_m) \quad (2-136)$$

具体到某一幅图像, 其所包含的灰度级  $M'$  往往为整个灰度级集合  $M$  的子集, 即  $M' \subseteq M$ 。同时, 某些灰度在一幅图像中出现的概率往往是很小的, 根据熵的可扩展特性, 这部分灰度对图像的整体信息熵的影响是很小的, 因此在计算图像信息熵时,

这部分灰度的影响可忽略不计。

## 2) 图像的单元熵

通过式(2-136)对图像信息熵的定义可知, 图像信息熵表现了图像灰度分布的全局统计特性, 图像信息熵的大小只与图像中各灰度出现的概率有关, 因此具有相同概率分布的两幅图像具有相同的信息熵。同时, 图像信息熵与图像的全局直方图一样仅仅考虑了图像的全局统计信息, 丢弃了图像的空间分布信息, 因此具有相同信息熵的图像可能在视觉上是完全不同的, 所以图像的全局信息熵不足以反映出图像间的差异。

我们知道, 不同图像之所以能呈现出不同的形状, 一方面是因为不同灰度的像素出现的频数不同, 另一方面也跟不同灰度的像素的空间分布特征有关。为了提取图像的形状特征, 在文献[90]中, 引入了网格描述符(Grid Descriptor, GD), 如图 2.32 所示, 在 GD 中, 图像首先被投射到具有固定分辨率的网格上, 若网格单元被图像内容覆盖(或被覆盖的程度超过一定阈值), 则该网格赋值为 1, 否则赋值为 0, 然后通过对该网格进行从左到右、从上到下的扫描, 形成一串二进制序列, 通过该二进制序列来描述图像的形状并进行图像间的相似性检索。

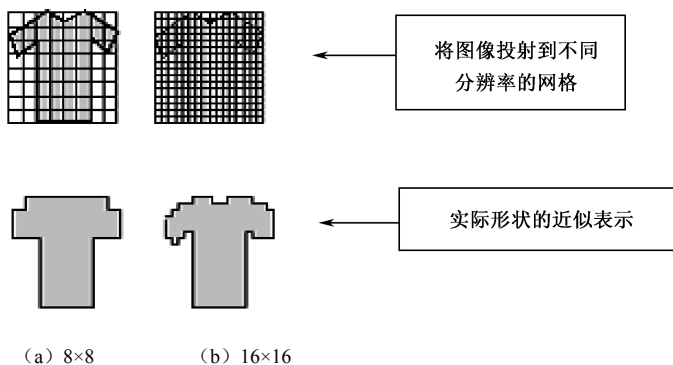


图 2.32 示例图像固定分辨率表示

但是在上述基于网格的描述方法中, 完全抛开了图像的灰度信息, 仅仅采用简单的赋 0、赋 1 来描述每一单元格的特征。基于此, 这里所提出的 EBIR 算法在基于 GD 方法的基础上, 采用图像局部灰度的统计特性来描述图像的形状信息。同 GD 方法一样, 首先将图像投射到具有固定分辨率的网格上(网格的分辨率随图像尺寸的变化而变化), 在对每一网格单元进行赋值时不是采用 GD 方法, 而是将包含该单元格灰度信息的信息熵作为该单元格的特征描述。为了描述图像单元格的这种特征, 引入单元熵的概念。设  $e_{ij}$  表示在图像空间中坐标为  $(i, j)$  的网格单元的单元熵, 则  $e_{ij}$  定义为

$$e_{ij} = - \sum_{m \in M} p'_m \log(p'_m) \quad (2-137)$$

其中,  $p'_m$  表示单元格内灰度为  $m$  的像素出现的概率。



采用单元熵对图像单元格的特征进行表示后,整幅图像就变成了一个由不同单元熵构成的熵矩阵,该熵矩阵包含了图像的全局和局部形状特征。采用熵矩阵来描述图像特征具有如下特点。

(1) 从全局图像来说,由于单元熵代表了图像局部单元的统计特性,因此由所有单元熵构成的熵矩阵不仅描述了图像的全局特性,也反映了图像的空间分布特性,这解决了采用图像全局熵描述图像特征所造成的熵相同而形状不同的问题,也解决了由于熵的对称性所造成的影响。

(2) 从局部上讲,由于各单元熵体现了图像局部灰度的统计特性,因此 EBIR 算法充分利用了图像的灰度信息来描述图像的特征,加强了图像检索的准确性。

(3) 利用熵矩阵描述形状,其准确度与所采用的单元格的分辨率有关,理论上分辨率越高对图像形状描述越准确,但在这种情况下,图像特征熵矩阵的维数会增大,存储图像特征所需的空间也会增大,同时还会造成图像检索速度的下降。

虽然图像的熵矩阵可以用来区分不同形状的图片,但如果直接以熵矩阵作为图片的形状特征用于图片的相似性检索,还存在下列问题。

(1) 维数高。很明显,所采用网格的分辨率与熵矩阵的维数密切相关。为了保证能尽量准确地体现原图片的形状特征,网格的分辨率不易太小,因此熵矩阵的维数将会很大,从而为图片的相似性度量带来困难。

(2) 在许多情况下,我们希望提取出的图片形状特征在平移、旋转、尺度等条件的变化下是一个不变量,显然,熵矩阵并不满足这种要求。

为此,在利用计算得到图片的熵矩阵后,我们采用不同的方法对熵矩阵进行处理,以克服上面提到的问题。

### 3) 利用熵矩阵的特征值向量进行检索

图片的特征采用熵矩阵描述后,即可采用熵矩阵来进行图片间的相似性度量,但采用熵矩阵作为图片的特征描述存在特征维数高的问题。考虑到图片中相邻的单元格之间一定的相关性,因此在熵矩阵中相邻的单元熵之间也具有一定的信息冗余。为了消除这种信息冗余,我们采用了熵矩阵的特征值向量来代替熵矩阵作为图片的索引特征,同时为了降低所提取的特征值向量的维数,我们采用特征值向量中模最大的几个特征值来进行图片间的相似性度量。对于特征值向量中模较小的特征值,由于其包含的信息量很少,因此忽略掉这些特征值不会对检索的结果造成太大的影响,而同时由于采用较少的特征值参与图片间的相似性计算,从而加快了图片检索的速度。

利用熵矩阵特征值向量进行检索的主要过程如下。

(1) 将图片投射到分辨率为  $n \times n$  的网格上,针对每一单元格,计算不同灰度级的概率分布向量,然后计算单元格的单元熵  $e_{ij}$ ,从而求取图片的熵矩阵  $\mathbf{E}$ 。根据上述熵的可扩展特性可知,如果图片单元格内的某一灰度的概率分布很小,则该灰度对单元格的信息熵的影响可忽略不计。因此,在计算图片的单元熵时可预先设定一个阈值  $\alpha$ ,

若  $p_m < \alpha$ ，在计算图像的单元熵时该灰度级可忽略不计。

$$\mathbf{E} = [e_{ij}] = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{bmatrix}, \quad 1 \leq i \leq n, 1 \leq j \leq n \quad (2-138)$$

(2) 将熵矩阵  $\mathbf{E}$  的特征值按照模的大小排列成一维向量  $(\lambda_1, \lambda_2, \dots, \lambda_n)$  (其中,  $\|\lambda_1\| \geq \|\lambda_2\| \geq \dots \geq \|\lambda_n\|$ ), 提取特征值向量的前  $t$  个主特征值作为图像的熵描述符。  $t$  取值满足下述条件, 即

$$\sum_{l=1}^t \lambda_l / \sum_{l=1}^n \lambda_l \geq \beta \quad (2-139)$$

其中,  $\beta$  为预定义阈值, 且  $\beta \in [0, 1]$ 。

(3) 对于提取的特征值向量, 利用欧氏距离对图像间的相似度进行度量, 即

$$d = \sqrt{\sum_{l=1}^t \|\lambda_l - \lambda_l^{(i)}\|^2} \quad (2-140)$$

其中,  $\lambda_l$  表示待检索图像的第  $l$  个特征值;  $\lambda_l^{(i)}$  表示用于检索的第  $i$  幅图像的第  $l$  个特征值。计算出图像间的距离后, 利用两幅图像间的距离判断出两幅图像的相似程度, 并依照事先设定的阈值检索出与例子图像相似的图像来。

#### 4) 利用熵矩阵的不变矩进行检索

不变矩是一个重要的基于区域的形状特征, 不变矩是通过所有属于区域内的像素点计算出来的, 因而不受噪声等因素的影响; 并且由于不变矩特征具有良好的尺度、平移和旋转不变性, 因而在基于形状的图像检索中取得了非常好的检索性能。这里我们在熵矩阵的基础上, 采用不变矩来描述图像特征<sup>[91]</sup>。

由上述方法得到的不变矩特征向量, 每一个元素的物理意义都是不同的, 因此它们的幅度也存在较大的差异, 如果直接用它们进行图像间的相似性度量, 会产生很大的偏差, 故必须通过归一化处理来消除这种偏差。

设图像的特征向量为  $\mathbf{R} = \{r_1, r_2, \dots, r_K\}$ ,  $K$  表示特征的维数, 特征内部归一化过程的目的在于使特征向量  $\mathbf{R}$  的各分量  $r_k$  ( $1 \leq k \leq K$ ) 具有同等的重要性。假设我们从数据库中选取  $M$  幅图像作为训练集, 设  $i$  为训练集中图像的索引值, 那么我们采用  $\mathbf{R}_i$  表示第  $i$  幅图像的特征向量, 即

$$\mathbf{R}_i = (r_{i,1}, r_{i,2}, \dots, r_{i,k}, \dots, r_{i,K}) \quad (2-141)$$

如果我们将所有的  $\mathbf{R}_i$  ( $1 \leq i \leq M$ ) 叠加在一起, 就可以得到维数为  $M \times K$  的矩阵, 即

$$\mathbf{V} = [v_{i,k}], \quad 1 \leq i \leq M, 1 \leq k \leq K \quad (2-142)$$

其中,  $v_{i,k}$  是特征向量  $\mathbf{R}_i$  (对应于第  $i$  幅图像) 的第  $k$  个分量。矩阵的第  $k$  列是维数为

$M$  的一个列向量, 记为  $\mathbf{V}_k$ 。我们的目的是将每列中的元素都归一化到一致的值域范围中, 以保证在计算两个向量的相似度时, 各分量具有同等重要性。

假设列向量  $\mathbf{V}_k$  是一个高斯数列, 我们首先计算该数列的均值  $\mu_k$  和标准差  $\sigma_k$ , 然后采用下式来进行归一化操作。

$$v'_{i,k} = \frac{v_{i,k} - \mu_k}{3\sigma_k} \quad (2-143)$$

根据  $3\sigma$  规则, 数列中各值落入区间  $[-1, 1]$  范围内的概率约为 99%。在实际应用中, 我们可以认为数列中的所有元素已落入区间  $[-1, 1]$  范围中了, 对于非此范围中的数值可以简单地对应到 -1 和 1 上。

## 2.3 纹理特征的提取与表达

### 2.3.1 概述

由于纹理基元及其分布形态复杂多样, 人们对纹理的感觉和心理效果相结合, 很难用语言文字来描述。尽管人们能很轻松地识别纹理, 但对纹理很难有一个确切的定义。一般将组成纹理的基本元素称为纹理基元或纹元。

Coggins 等收集了计算机视觉领域中一些经典的纹理定义<sup>[92]</sup>。

(1) 纹理可以被认为是由肉眼可见的区域组成的。纹理结构的简单特征是有重复图案的组成, 在这些图案中的图元按一定的布局规则排列。

(2) 如果图像的一组局部统计特征或者其他特征是不变的、变化缓慢的或者近似周期的, 那么就认为图像区域含有不变的纹理。

Castleman 等人认为<sup>[93]</sup>: 纹理是一种反映图像中一块区域的像素灰度级的空间分布属性, 这种空间结构的固有属性可以通过邻域像素间的相关性刻画。

以上对纹理的描述已慢慢地被广大学者接受和应用。对纹理的认识或定义决定了纹理特征提取采用的方法。由于对纹理的定义不统一, 一方面使纹理分析中的问题更为复杂、更具有挑战性; 另一方面, 由于纹理本身具有多种属性使得图像的研究者们引入各种模型对纹理特征进行描述, 从而使纹理的研究丰富多彩。

较为常见的纹理主要有以下 3 种类型。

(1) 自然纹理。该种纹理是未经人工刻意加工的, 是在自然界中自然存在的物体表面属性, 如烟、雾、白云、木头、砾岩、沙漠、草地纹理。这种纹理的基本组成元素形状多样, 多数不规则, 分布随机性较大。

(2) 人工纹理。该种纹理是人工制造的, 是不同于自然存在物体表面属性的一种纹理, 如器物表面的花纹、砖墙、织物、棋盘格等。这种纹理的主要特点是纹理基本

组成元素形状规则确定,分布规律性比较强。

(3) 混合纹理。这种纹理主要是一些人工制造的纹理基本元素随机分布于物体表面或自然界中而形成的。

纹理特征是指利用计算机技术从数字图像中计算出来的、可以定量描述人对纹理的定性感知的某些参数,它对区域内部灰度变化或色彩变化的某种规律进行量化。这些纹理特征能够尽可能地缩小纹理的类内差距,同时尽可能增大纹理的类间差距。纹理的视觉特征一般有 3 个基本量:周期性、方向性和随机性。其中,周期性和方向性是两个高层次的纹理特征,可以用来指导纹理图像的知觉感知。不同的应用问题和不同的图像类型都给图像纹理特征提出了不同的需求。

### 2.3.2 常用的纹理分析方法

纹理分析指的是通过一定的图像处理技术提取纹理特征,并获得纹理定性或定量描述的过程。纹理分析包括两大部分:检测纹理基元、获得相关纹理基元排列分布方式的信息<sup>[94]</sup>。常用的纹理分析方法有 4 种:统计分析方法、结构分析方法、模型分析方法和频谱分析方法。

#### 1. 统计分析方法

纹理特征,特别是自然纹理,在局部上表现出很大的随机性,可描述成一个随机变量。但从整体和统计意义上看,它也存在某种规律性。从区域统计方面去分析纹理图像的方法称为基于统计的分析方法,该方法是利用图像空间灰度分布情况来描述粗细度、均匀性、方向性等纹理信息。

较早提出并应用的一种统计方法是利用自相关函数描述图像的纹理特征。1973 年,Haralick 等人<sup>[95]</sup>提出了空间灰度共生矩阵法,该方法首先对图像空间灰度分布进行统计,得出图像的共生矩阵,其次依据定义对共生矩阵上的若干个纹理特征值进行计算,得到图像的纹理描述。由于共生矩阵模型方法不受分析对象的制约,能够很好地反映图像空间灰度分布情况,体现图像的纹理特征,所以得到广泛应用。在此基础上,洪继光等结合图像灰度信息及灰度变化的梯度信息,提出了灰度-梯度共生矩阵法<sup>[96]</sup>。该方法描述图像的特征除了利用灰度本身之外,还利用灰度变化的梯度信息。图像灰度大小构成了图像的基础,图像梯度则构成了图像轮廓、边缘的要素。我们将结构分析方法和统计分析方法相结合,并以方块编码为依据,提出了一种纹理基元的共生矩阵方法<sup>[97]</sup>。近年来,灰度共生矩阵及其改进方法仍然在纹理分析及模式识别领域广泛应用<sup>[98,99]</sup>。

为了满足人类对纹理的视觉感知心理学的研究,1978 年,Tamura 等人<sup>[100]</sup>提出了用纹理的 6 种视觉特征来表示纹理,这种表示纹理的方法使表示的纹理性质具有直观

的视觉意义。该纹理特征和灰度共生矩阵的主要区别是 Tamura 纹理特征中所有纹理特性都在视觉上有意义。Tamura 纹理特征包括 6 个分量, 对应心理学角度上纹理特征的 6 种属性, 分别是粗糙度 (coarseness)、对比度 (contrast)、方向度 (directionality)、线性度 (linearity)、规整度 (regularity) 和粗略度 (roughness)。为描述中心像素与周围邻域像素之间的相对灰阶关系, 盛文等<sup>[101]</sup>于 2000 年提出了一种基于纹理元灰度模式统计的图像纹理分析方法, 与其他方法相比, 该方法方便简单, 计算量较少, 得到了越来越广泛的应用。随着研究深入, 多种基于统计法的纹理分析方法被提了出来, 如 Xie 等<sup>[102]</sup>提出的 TEXEM 模型、Quan 等<sup>[103]</sup>提出的 pattern fractal spectrum 方法等。

近年来, 一种简单有效的纹理分析方法——局部二值模式 (Local Binary Pattern, LBP) 得到了广泛研究。由于 LBP 原理相对简单, 计算复杂度低, 同时融合了纹理的结构特征和统计特征, 且不受光照变化等因素的影响, 因而在纹理分析、人脸识别、运动分析、医学图像分析等领域得到了广泛应用。后面将主要针对 LBP 进行详细讨论。

## 2. 结构分析方法

结构法纹理分析的基本思想是假定复杂的纹理模式是由简单的纹理基元 (基本纹理元素) 以一定的有规律的形式重复排列组合而成的。结构分析方法适用于印刷图像, 如布料、砖墙类等人工形成的纹理图像, 其纹理基元和排列比较规则, 按纹理基元的特性和其排列规则来描述。

1966 年, Beck<sup>[104]</sup>以不同的英文字母作为纹理基元进行观察, 发现纹理基元按不同方向分布影响着人们对纹理的区分。在此基础上, Bergen 等<sup>[105]</sup>发现纹理基元的方向和纹理基元的密度都显著影响着人们对不同纹理的区分, 同时, 纹理基元的大小及尺寸之间的对比, 也对纹理的区分有着重要的影响。从而从生理和心理的角度说明纹理图像可以分解为纹理基元, 而结构分析方法就是按纹理基元的特性和其排列规则来描述的。

比较规则的纹理在空间中以有次序的形式进行纹理单元的镶嵌, 最典型的模式是用一种正多边形镶嵌而成, 如 Voronoi 多边形<sup>[106]</sup>。文献[107]提出了图状语法结构定义排列规则的纹理模型, 该模型使用直线段、开放多边形和封闭多边形作为纹理基元。由于纹理结构的复杂性, 图状语法结构比较简单, 文献[108]提出了树型语法结构表示纹理, 将纹理按照  $9 \times 9$  的窗口进行分割, 每个分解单元的空间结构表示一棵树。

结构分析方法的好处是纹理构成容易理解, 适合于高层检索、描述规则的人工纹理。但对不规则的自然纹理, 由于基元本身提取困难及基元之间的排布规则复杂, 因此结构法受到很大的限制。

## 3. 模型分析方法

基于模型的纹理分析是以图像的构造模型为基础, 通过模型参数来定义纹理, 模

型的参数决定纹理的质量。模型法的主要问题是估计模型参数,使其所表示的纹理图像逼近原纹理图像。典型的模型有随机场模型、分形模型、Wold 模型等。

常见的随机场模型有马尔可夫模型、Gibbs 模型等。马尔可夫随机场(Markov Random Field, MRF)模型<sup>[109]</sup>的纹理分析方法把纹理看作一个随机的二维图像场,并且假定某一点取值与周围像素取值多少有关。近年来,MRF 模型在纹理分析中得到了广泛应用<sup>[110,111]</sup>。但 MRF 模型仅通过局部特征很难得到全局的联合分布,于是 Geman 等提出了 Gibbs 随机场模型<sup>[112]</sup>,该模型通过集团势能的概念,利用局部的计算获得全局的结果。同时针对 MRF 的各种方法也不断出现,如 Cohen 等<sup>[113]</sup>提出了高斯-马尔可夫模型。MRF 的另一种应用实例是自回归纹理(Simultaneous Auto-Regressive, SAR)模型。在 SAR 模型中,每个像素的强度被看成随机变量,可以通过其相邻的像素来描述。另外,SAR 模型的一种变化称为旋转无关的自回归纹理模型<sup>[114]</sup>,它具有和图像的旋转无关的特点。但定义合适的 SAR 模型需要确定相邻像素集合的范围,而固定大小的相邻像素集合范围不能很好地表达各种纹理特征。为此,Luo 等提出了多维度的自回归纹理模型<sup>[115]</sup>,该模型能够在多个不同的相邻像素集合范围下计算纹理特征。Bennett 等人<sup>[116]</sup>在分析比较 MRF 模型和 SAR 模型的基础上,提出一种广义长相关模型,该模型可以对具有长相关性质的低频纹理图像进行很好的建模。

对于图像中分形的研究,较早的研究认为大多数图像纹理表面的粗糙度与分形模型很接近,自然界的物体(如云雾、青山、凹凸不平的地面、风化而斑驳的岩石)大多具有比较强的分形特征,分形模型在一定的尺度范围内可以很好地与自然背景的表面和实际结构相吻合。这就使我们可以利用各种不同的特性参数(分形特性)来区分不同的物体。分形维数是分形对象极其重要的特征,它可以对分形对象的复杂度、不规则程度和全局正则性进行定量描述。分形维数在图像处理中的应用是以两点为基础的:一是自然界中不同种类的形态物质一般具有不同的分形维数;二是自然界中的分形与图像的灰度表示之间有着一定的对应关系。准确地估计分形对象的分形维数在分形信号的处理和分析中非常关键。目前常用的分形维数有相似维数、Hausdorff 维数、计盒维数等<sup>[117,118]</sup>。

基于 Wold 模型的随机场分割法则将图像随机场分解为确定性和不确定性两种类型,能更精确地刻画纹理的特点。Wold 方法最早被用于 PHOTOBOOK 系统中,它的基本原理是随机场分解 Wold 原理,即二维随机场可以被分解为 3 个正交的元素:谐波周期性分量、渐进为零的方向分量和一个非确定性的随机分量。Wold 分解首先检测图像的谐波分量,并提取谐波峰值特征,再对剔除谐波分量后的随机场进行二阶多分辨率自回归建模,提取模型参数和对应的协方差矩阵作为纹理特征。基于 Wold 模型的纹理表示方法在纹理分析领域也得到广泛应用<sup>[119,120]</sup>。

#### 4. 频谱分析方法

频谱法主要借助各种变换算法利用图像的频率特性来描述纹理特征。其关键主要是寻求一种可逆的线性变换，从看似复杂的数据中找出一些直观的信息，再对它进行分析，从而可以用一组不相关的数据（通常是一组系数）来代替图像数据。将这些系数按其含有图像信息及对图像主观质量影响的重要程度排序，删除一些不会对图像内容描述产生重大影响的不重要系数，用少量的高效的系数进行图像的特征描述。由于图像信号往往在频域具有比在空域更加简单和直观的特性，所以频谱分析法在图像分析中起着很重要的作用。常用的方法有傅里叶变换法、小波变换法、Gabor 变换法等。

傅里叶变换在纹理图像的分析中具有许多优点，如具有明确的物理意义、其功率谱具有平移不变性、有快速算法等。因此，在纹理分析中，常借助傅里叶频谱的频率特性来描述周期的或近乎周期的二维图像模式的方向性<sup>[121]</sup>。小波变换对图像边界和奇异点的检测十分有效，所以特别适用于图像纹理特征的提取<sup>[122]</sup>。针对小波分析技术在纹理特征提取方面的应用，近年来出现了很多研究成果<sup>[123,124]</sup>。1946 年，Gabor 博士针对傅里叶变换存在不能同时进行时间、频率局部分析的缺点，提出了 Gabor 函数。生物学研究发现，二维 Gabor 滤波器可以模拟生物的视觉系统，能够很好地描述脊椎动物大脑初级视觉皮层部分的单细胞可接受信息域的分布，因而在图像分析中具有重要的作用。而且，Gabor 滤波器在消除空域和频域二维联合不确定性方面是最优的，它可以看成是方向、尺度可调的边界和直线检测器，是被公认的信号表示尤其是图像辨识的最好方法之一。所以，可以通过 Gabor 滤波器检测出图像中不同方向和角度上的边缘和线条，以提取图像中的纹理特征。Manjunath B.S.等人<sup>[125]</sup>根据 Gabor 滤波函数是完备的非正交函数集，消除系列 Gabor 滤波器的冗余度（相关性），同时又设计了一种自适应滤波器选择方法，从而使其可有效描述纹理特征。此后，多种基于 Gabor 变换的纹理描述方法被提了出来<sup>[126-128]</sup>。

#### 5. 几种常用纹理特征

##### 1) 灰度共生矩阵

由于纹理是相邻像素或相邻小区域灰度上及几何位置等相互关系的表征，因此处于同样位置关系的一对像素的某种条件概率就可以用来表示其纹理特征。Haralick 等人从数学角度研究了图像纹理中灰度级的空间依赖关系，根据图像中各像素之间的角度方位和距离关系构造了一个灰度共生矩阵（Gray Level Co-occurrence Matrix, GLCM）<sup>[95]</sup>。该矩阵按照图像灰度值的空间关系描述像素点对之间的空间结构特征及其相关性，表示相距  $(\Delta x, \Delta y)$  的两个灰度像素同时出现的联合分布概率。如图 2.33 所示，其中， $\Delta x$  和  $\Delta y$  的范围由像素间距  $\delta$  和方向  $\theta$  两个参数决定，即  $\Delta x = \delta \cos \theta$ ，

$\Delta y = \delta \sin \theta$ 。该方法可以表示纹理的稀疏度、对比度、复杂度及其纹理力度等特性。

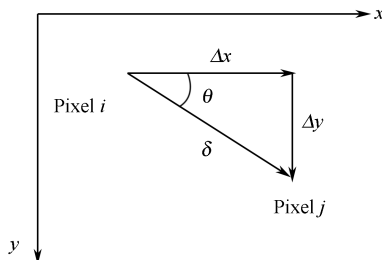


图 2.33 像素对分布示意图

设  $f(x, y)$  表示一幅灰度图像，对图像中的任一区域  $R$ ，定义  $S$  为区域中具有特定空间联系的像素对的集合，则灰度共生矩阵的元素可表示为

$$m_{(d,\theta)}(i, j) = \text{card} \{ [(x_1, y_1), (x_2, y_2)] \in S \mid f(x_1, y_1) = i, f(x_2, y_2) = j \} \quad (2-144)$$

其中， $x_2 = x_1 + d \cos \theta$ ； $y_2 = y_1 + d \sin \theta$ 。 $\text{card}(S)$  表示集合  $S$  中对  $m_{(d,\theta)}(i, j)$  有贡献的元素个数。在实际应用中，常需要对式 (2-144) 进行归一化，得

$$m_{(d,\theta)}(i, j)' = \frac{\text{card} \{ [(x_1, y_1), (x_2, y_2)] \in S \mid f(x_1, y_1) = i, f(x_2, y_2) = j \}}{\text{card}(S)} \quad (2-145)$$

为减少计算量，计算中往往需要先对图像进行灰度变换，降低灰度级，同时减少  $\theta$  的方向数，通常取  $0^\circ$ 、 $45^\circ$ 、 $90^\circ$ 、 $135^\circ$  四个方向。

共生矩阵体现了不同纹理之间的区别，对于具有不同特点的图像纹理，其共生矩阵明显不同。灰度共生矩阵中的主对角线上元素是一定位置关系下的两像素相同灰度值组合出现的次数。由于存在沿纹理方向上相近像素的灰度值基本相同，垂直纹理方向上相近像素的灰度值差异较大的一般规律，因此，这些主对角线元素的大小有助于判别纹理的方向和粗细，对纹理分析起着重要的作用。对于纹理较为粗糙的图像，像素对一般具有相同的灰度，其灰度共生矩阵中的值主要集中于主对角线附近；而对于纹理较为细腻的图像，由于其像素对灰度差异较大，其灰度共生矩阵中的值则散布在各处。例如， $m_{(1,0)}(i, j)$  描述一幅图像水平方向上的灰度级的联合分布，如果这个共生矩阵的主对角线上的元素全部为零，则说明水平方向没有相邻的两个像素具有相同的值，即水平方向上灰度变化频繁，纹理较为细腻；如果主对角线上的元素很大，则表明水平方向上纹理较为粗糙。

灰度共生矩阵把像素作为纹理基元，表示基元的特征是像素位置和灰度，基元之间的空间关系也被局限于空间位移向量，虽然在一定程度上能够很好地体现图像中的纹理信息，但在某些情况下对图像的特征描述常常需要对灰度共生矩阵进行推广。纹理基元可选用图像中的低层单元，如边缘点、边缘段或灰度均匀的小区域等。描述这些纹理基元的特征可以是多个。例如，边缘单元可用位置、对比度、方向、灰度均值、灰度方差、区域尺寸和形状等信息来描述。同样，纹理基元之间的空间关系可以推广



为用任意复杂的空间约束来定义的通用关系。

设  $U$  为图像中所有基元的集合,  $V$  为图像中基元性质的集合,  $f$  为赋予  $U$  中基元一个  $V$  中性质的函数。在基元间利用距离或相邻性建立空间联系, 设  $S \subseteq U \times U$  为满足上述空间联系的基元对间的联系集合, 则推广的共生矩阵  $\mathbf{M}$  的元素定义为

$$m(v_1, v_2) = \frac{\text{card}\{(u_1, u_2) \in S | f(u_1) = v_1, f(u_2) = v_2\}}{\text{card}(S)} \quad (2-146)$$

其中,  $m(v_1, v_2)$  为满足上述空间联系的基元对 (一个基元具有性质  $v_1$ , 而另一个基元具有性质  $v_2$ ) 的相对频率;  $\text{card}(S)$  的定义同式 (2-145)。

灰度共生矩阵是分析图像局部模式和排列规则的基础, 为了有效利用灰度共生矩阵所提供的图像灰度方向、间隔和变化幅度的信息, 可以在共生矩阵的基础上提取一些有意义的统计量, 基于这些数字统计量, 对图像的纹理特征进行定量描述。所提取的纹理特征主要有以下几种。

(1) 角二阶矩 (能量)。该特征反映了图像区域的均匀性或平滑性。

$$W_1 = \sum_i \sum_j [m(i, j)]^2 \quad (2-147)$$

在均匀区域, 灰度值变化较小, 大部分像素对具有相同或相近的取值, 主要发生在矩阵的对角线附近, 其他大部分元素为零; 而在非均匀区, 灰度值变化大的像素对较多, 在整个灰度共生矩阵上概率均匀分布, 而且矩阵中元素的值都很小。所以, 非均匀区的角二阶矩比均匀区域的角二阶矩要小。该测度对区域内部有无灰度值变化较敏感, 但对灰度值变化数值大小不敏感, 即具有高的局部灰度值对比度的区域角二阶矩值不一定高。当共生矩阵中所有  $m(i, j)$  都相等时,  $W_1$  达到最小值。

(2) 对比度。对比度又称非相似性, 可理解为图像的清晰度, 即纹理清晰度, 可表示为

$$W_2 = \sum_i \sum_j (i - j)^2 m(i, j) \quad (2-148)$$

其中,  $(i - j)$  表示图像特定位置关系下像素对的灰度值差, 灰度值差大的像素对越多, 这个值就越大。对比度反映了近邻像素的反差, 当图像中两个灰度级点对的统计个数接近共生矩阵对角线时, 纹理变化较小, 对比度较小; 反之, 则表明近邻像素的反差较大, 纹理较细。因此, 对比度值的大小反映了纹理的粗细度。

(3) 相关系数。相关系数在一定程度上反映了矩阵行与列的线性相关程度。相关系数较大时图像区域灰度分布比较均匀。

$$W_3 = \frac{\sum_i \sum_j ijm(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (2-149)$$

其中,  $\mu_x$ 、 $\mu_y$ 、 $\sigma_x$ 、 $\sigma_y$  分别定义为

$$\begin{aligned}
 \mu_x &= \sum_i i \sum_j m(i, j) \\
 \mu_y &= \sum_j j \sum_i m(i, j) \\
 \sigma_x^2 &= \sum_i (i - \mu_x)^2 \sum_j m(i, j) \\
 \sigma_y^2 &= \sum_j (j - \mu_y)^2 \sum_i m(i, j)
 \end{aligned} \tag{2-150}$$

(4) 熵。熵给出了图像内容随机性的度量，是图像具有信息量的度量，可表示为

$$W_4 = - \sum_i \sum_j m(i, j) \log_2 [m(i, j)] \tag{2-151}$$

纹理信息是图像信息的一种，若图像没有任何纹理，则灰度共生矩阵几乎为零矩阵，熵值接近于零；若图像有较多的细纹理，则矩阵中元素近似相等，该图像的熵值最大；若仅有较少的纹理，则矩阵中的元素差别较大，图像的熵值就较小。

(5) 差分矩。差分矩的定义为

$$W_5 = \sum_i \sum_j (i - \mu)^2 m(i, j) \tag{2-152}$$

其中， $\mu$  为  $m(i, j)$  的均值。

(6) 逆差分矩。逆差分矩又称均匀性，定义为

$$W_6 = \sum_i \sum_j \frac{m(i, j)}{1 + (i - j)^2} \tag{2-153}$$

对于匀质区域，其灰度共生矩阵的元素集中在对角线上， $(i - j)^2$  值小，则均匀性的值较大；对非匀质区域，由于其灰度共生矩阵的元素集中在远离对角线上， $(i - j)^2$  值大，则均匀性的值较小。所以，该特征是图像分布平滑性的测度。

(7) 和平均。和平均的定义为

$$W_7 = \sum_k \sum_i \sum_j i m(i, j) \tag{2-154}$$

其中， $k = i + j$ 。

(8) 和方差。和方差的定义为

$$W_8 = \sum_k \sum_i \sum_j (i - W_7)^2 m(i, j) \tag{2-155}$$

其中， $k = i + j$ 。

(9) 和熵。和熵的定义为

$$W_9 = - \sum_k \sum_i \sum_j m(i, j) \log_2 m(i, j) \tag{2-156}$$

(10) 差方差。差方差的定义为

$$W_{10} = \sum_{k=2}^{2N} \sum_{i=1}^N \sum_{j=1}^N m(i, j), \quad k = |i - j| = 0, 1, \dots, N - 1 \tag{2-157}$$

其中,  $d = \sum_{i=1}^N \sum_{j=1}^N |i - j| m(i, j)$ 。

在共生矩阵基础上定义的纹理描述符能够对图像特征进行很好的描述, 但其最大的缺陷在于这些统计特征纯粹从数学角度推算而来, 和人们在视觉上对纹理特征的鉴别和感知无法建立对应关系; 同时, 由于灰度共生矩阵本身具有方向性, 从矩阵中提取的统计量也只能反映某一方向信息, 虽然可以采用几个方向的统计量作平均, 但还是无法很好地表达图像的空间信息, 对于包含不同纹理区域的图像, 无法体现图像中各纹理区域之间的空间位置关系; 而且从矩阵中提取的特征向量还不具备尺度不变性。

## 2) Tamura 纹理特征

基于人类对纹理的视觉感知研究, Tamura 等人提出了一种纹理特征的表达<sup>[100]</sup>。该纹理特征和灰度共生矩阵的主要区别是 Tamura 纹理特征中所有纹理特性都在视觉上有意义, 而灰度共生矩阵的某些纹理属性不具有视觉意义 (如信息熵), 这一特点使得 Tamura 纹理特征在图像检索中使用较多。Tamura 纹理特征包括 6 个分量, 对应心理学角度上纹理特征的 6 种属性, 分别是粗糙度 (coarseness)、对比度 (contrast)、方向度 (directionality)、线性度 (linearity)、规整度 (regularity) 和粗略度 (roughness)。其中, 前 3 个分量对图像检索尤为重要, 下面着重讨论这 3 种特征的定义和数学表达。

### (1) 粗糙度

粗糙度是反映纹理中粒度的一个量, 是最基本的纹理特征, 因此, 从狭义的观点来看, 纹理就是粗糙度。当两种纹理模式只是基元尺寸不同时, 具有较大基元尺寸的模式给人的感觉更粗糙。而对具有不同结构的纹理模式来说, 基元尺寸越大或者基元重复次数越少, 给人的感觉越粗糙。粗糙度可由不同大小窗口的像素的滑动均值得到, 计算可以分以下几个步骤进行。

首先, 计算图像中大小为  $2^k \times 2^k$  个像素的活动窗口中像素的平均强度值, 即有

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i, j) / 2^{2k} \quad (2-158)$$

其中,  $k = 0, 1, \dots, 5$ ;  $g(i, j)$  是位于  $(i, j)$  处像素的灰度值。

然后, 对于每个像素, 分别计算它在水平和垂直方向上互不重叠的窗口之间的平均强度差, 即

$$\begin{aligned} E_{k,h}(x, y) &= |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \\ E_{k,v}(x, y) &= |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})| \end{aligned} \quad (2-159)$$

取水平和垂直两个方向上的最大均值差值为当前像素的邻域均值差值, 即

$$E_k(x, y) = \max[E_{k,h}(x, y), E_{k,v}(x, y)] \quad (2-160)$$

对于每个像素, 从多邻域尺寸中设置一个最佳尺寸, 即

$$S_{\text{best}}(x, y) = 2^k + 1 \quad (2-161)$$

使得  $E_k = E_{\max} = \max(E_1, E_2, \dots, E_L)$ 。其中,  $E$  为邻域均值差值;  $L$  为邻域尺寸个数。

最后, 计算整幅图像中  $S_{\text{best}}$  的平均值作为灰度图像的纹理粗糙度, 表示为

$$F_{\text{crs}} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{\text{best}}(i, j) \quad (2-162)$$

其中,  $m$  和  $n$  分别是图像的宽度和高度。

显然, 整幅图像的粗糙度可以在一定程度上反映纹理图像的粒度特性, 但存在计算量大、衡量不准确等缺点。粗糙度特征的另一种改进形式是采用直方图来描述  $S_{\text{best}}$  的分布, 而不是如上述方法那样简单地计算  $S_{\text{best}}$  的平均值。这种改进后的粗糙度特征能够表达具有不同纹理特征的图像或区域, 对图像检索更为有利。

### (2) 对比度

对比度是反映像素亮度统计分布的一个量, 其大小由以下 4 个因素决定: 灰度级动态范围、直方图上黑白两部分两极化的程度、边缘的锐度、重复模式的周期。一般意义上, 我们说纹理的对比度是指前两个因素, 可以通过对像素强度分布情况的统计得到, 确切地说, 它是通过  $\alpha_4 = \mu_4 / \sigma^4$  来定义的, 其中,  $\mu_4$  是四次矩,  $\sigma^2$  是方差。对比度通过如下公式衡量。

$$F_{\text{con}} = \frac{\sigma}{(\alpha_4)^n} \quad (2-163)$$

该值给出了图像或区域中对比度的全局变量,  $n$  通常取值为 8、4、2、1、1/2、1/4 或 1/8。

### (3) 方向度

方向度指的是给定纹理区域的全局特性, 描述了纹理是如何沿某些方向散布或集中的。一般来说, 方向度与纹理基元的形状及如何将这些纹理基元排列的规则有关, 它是基于像素的梯度向量计算的。计算方向度时, 首先计算每个像素处的梯度向量, 该向量的模和方向分别定义为

$$\begin{aligned} |\nabla G| &= (|\Delta_H| + |\Delta_V|) / 2 \\ \theta &= \arctan(\Delta_V / \Delta_H) + \pi / 2 \end{aligned} \quad (2-164)$$

其中,  $\Delta_H$  和  $\Delta_V$  分别是通过图像卷积下列两个  $3 \times 3$  操作符所得的水平和垂直方向上的变化量。

$$\begin{array}{ccc} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{array} \quad \begin{array}{ccc} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{array} \quad (2-165)$$

Tamura 方法的核心就是构建方向角局部边缘概率直方图。当所有像素的梯度向量被计算出来后, 对  $\theta$  值范围进行离散化, 统计每个柄中相应的  $|\nabla G|$  大于给定域值的像素数量, 构建方向角局部边缘概率直方图。该直方图对于具有明显方向性的图像会表现出峰值, 对于无明显方向性的图像则表现比较平坦。可以根据方向角局部边缘概率直方图计算直方图波峰尖锐程度来定量地描述纹理的方向性。Tamura 等人采用二阶矩

累加的方法，其定义为

$$F_{\text{dir}} = 1 - rn_p \sum_p \sum_{\phi \in \omega_p}^{n_p} (\phi - \phi_p)^2 H(\phi) \quad (2-166)$$

其中， $p$  代表直方图的峰值； $n_p$  为直方图中波峰的个数； $\omega_p$  代表峰值两侧谷底距离； $\phi_p$  是波峰中心位置； $r$  为直方图归一化系数； $\phi$  为量化后的方向角； $H$  为直方图。

### 3) Gabor 变换法

由于 Gabor 变换能够很好地同时在时域和频域中兼顾对信号分析的分辨率的要求，因此该变换成为最常用的图像纹理特征提取方法。

Gabor 函数由 Gabor 博士于 1946 年提出，通过高斯函数加上频移后产生。Gabor 滤波器是用 Gabor 函数作为单位冲激响应的带通滤波器，具有良好的滤波性能，其输出可以看作输入信号的 Gabor 小波变换。生物学研究发现，二维 Gabor 滤波器可以模拟生物的视觉系统，能够很好地描述脊椎动物大脑初级视觉皮层部分的单细胞可接受信息域的分布，因而在图像分析中具有重要的作用。而且，Gabor 滤波器在消除空域和频域二维联合不确定性方面是最优的，它可以看成是方向、尺度可调的边界和直线检测器，是被公认的信号表示尤其是图像辨识的最好方法之一。所以，可以通过 Gabor 滤波器检测出图像中不同方向和角度上的边缘和线条，以提取图像中的纹理特征。

由于 Gabor 滤波可以看作一种小波变换，因此可以从小波变换的角度阐述该滤波器组的定义。

设图像为  $f(x, y)$ ，它的二维小波变换为

$$I_{mlpq} = \iint f(x, y) \varphi_{ml}(x - p\Delta x, y - q\Delta y) dx dy \quad (2-167)$$

其中， $\Delta x$  和  $\Delta y$  是空间采样间隔，通常设  $\Delta x = \Delta y = 1$ ； $p$  和  $q$  是图像像素点的位置； $m$  和  $l$  分别定义了小波变换的方向和尺度，即  $m = 0, \dots, M-1$ ， $l = 0, \dots, L-1$ ； $\varphi_{ml}(x, y)$  由小波变换的母波得到，即

$$\varphi_{ml}(x, y) = a^{-m} \varphi(\tilde{x}, \tilde{y}) \quad (2-168)$$

其中， $\tilde{x} = a^{-m}(x \cos \theta + y \sin \theta)$ ； $\tilde{y} = a^{-m}(-x \sin \theta + y \cos \theta)$ 。

母波  $\varphi(x, y)$  通过  $a^{-m}$  尺度发生变化，方向  $\theta$  变化的定义为

$$\begin{aligned} \theta &= l \cdot \Delta \theta \\ \Delta \theta &= 2\pi / L \end{aligned} \quad (2-169)$$

这个定义使得所有的滤波器具有同样的能量。当我们把 Gabor 函数作为母小波时，Gabor 滤波就可以看成小波变换。

$$g(x, y) = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W_x \right] \quad (2-170)$$

其中， $W$  定义了滤波器的频率带宽。由神经物理学的研究， $W = 0.5$ ，即取数字最高频率 1 的一半时，完全符合人类的视觉系统。

由 Gabor 小波的定义我们可以看出, Gabor 小波实质上是由一个复指数信号调制一个二维高斯函数得到的。设频域空间的坐标轴是  $u$  轴和  $v$  轴, Gabor 小波的频域特征是由高斯函数的频谱沿着  $u$  轴平移得到的, 即

$$\begin{aligned}\psi(u, v) &= e^{-2\pi^2(\sigma_u^2 u^2 + \sigma_v^2 v^2)} * \delta(u - W) \\ &= e^{\left\{-2\pi^2\left[\sigma_x^2(u-W)^2 + \sigma_y^2 v^2\right]\right\}} \\ &= e^{\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\}}\end{aligned}\quad (2-171)$$

其中, 频率方差  $\sigma_u = \frac{1}{2\pi\sigma_x}$ ;  $\sigma_v = \frac{1}{2\pi\sigma_y}$ 。

在纹理分析中, 滤波器组的各个滤波器具有相同的能量, 但是在计算自然图像的相似性时, 由于自然图像的频谱幅度由低频到高频迅速下降, 导致 Gabor 滤波器的高频输出非常小。因此, 我们在计算小波变换的各个基函数时, 重新定义式 (2-168), 得

$$\hat{\phi}_{ml}(x, y) = \phi(\tilde{x}, \tilde{y}) \quad (2-172)$$

这样, 式 (2-167) 可以重新定义为

$$\hat{I}_{mlpq} = \iint f(x, y) \hat{\phi}_{ml}(x - p\Delta x, y - q\Delta y) dx dy \quad (2-173)$$

设图像的每一个像素点  $(p, q)$  经过滤波后, 得到的输出特征为

$$F_{mlpq} = \left| \hat{I}_{mlpq} \right|, \quad m = 0, \dots, M-1, \quad l = 0, \dots, L-1 \quad (2-174)$$

输出特征值保留了能量信息, 忽略了位置信息。在提取图像的纹理特征时, 能量信息可以较好地反映图像的特征。

这样给定一幅图像后, 它的纹理特征可以由 Gabor 滤波器的输出得到, 即

$$E_{ml} = \frac{\sum_{p,q} F_{mlpq}^2}{\sum_{m,l,p,q} F_{mlpq}^2} \quad (2-175)$$

上式中的分母是标准化因子, 它的作用是增强图像内容, 减少图像在拍摄过程中光照度和对比度对图像特征的影响。

#### 4) MRF 模型法

近 20 年来, 由于 MRF 模型在理论上可以产生任何模式的纹理, 所以利用 MRF 模型法来描述纹理特征取得了很大的成功。这一模型假设每个像元的密度与邻域像元有关, 与其他像元无关, 紧靠的元素有直接交互作用, 另外, 全局的影响也可以传播。在纹理分析中, MRF 模型基于纹理满足随机、静态等前提条件。采用 MRF 模型来描述纹理图像时, 首先需要对图像进行分块, 在每个分块中采用最大似然估计 (为计算方便经常采用伪似然估计) 和最小二乘估计等方法估计模型参数, 然后对一系列的模

型参数进行聚类, 确定该像素点及其邻域情况下该像素点最可能归属的概率。

对于一幅给定的图像  $\{y_0(i, j)\}$ , 令  $y = y_0 - \mu$ ,  $\mu$  为  $y_0$  的均值, 则零均值输出  $\{y(i, j)\}$  满足下列差分方程。

$$y(i, j) = \sum_{(k, l) \in \eta} g_{k, l} y(i - k, j - l) + e(i, j) \quad (2-176)$$

其中,  $\{g_{k, l} | (k, l) \in \eta\}$  为 MRF 模型参数, 在这里为待估量,  $\eta$  为  $(i, j)$  的邻域(指标集);  $\{e(i, j)\}$  为零均值的高斯噪声场, 其相关结构为

$$E[e(i, j)e(m, n)] = \begin{cases} \sigma_e^2, & (i, j) = (m, n) \\ -g_{i-m, j-n} \sigma_e^2, & (m, n) \in \eta \\ 0, & \text{其他} \end{cases} \quad (2-177)$$

$$E[y(i, j)e(m, n)] = \begin{cases} \sigma_e^2, & (i, j) = (m, n) \\ 0, & \text{其他} \end{cases}$$

对 MRF, 有  $g_{k, l} = g_{-k, -l}$ , 从而 MRF 纹理模型可表示为

$$y(i, j) = \sum_{(k, l) \in \eta_s} g_{k, l} [y(i - k, j - l) + y(i + k, j + l)] + e(i, j) \quad (2-178)$$

其中,  $\eta_s$  为非对称域, 它满足

$$\left. \begin{aligned} \eta &= \eta_s \cup \bar{\eta}_s \\ \eta_s \cap \bar{\eta}_s &= \Phi \\ \bar{\eta}_s &= \{(-k, -l) | (k, l) \in \eta_s\} \end{aligned} \right\} \quad (2-179)$$

记点  $(i, j)$  为  $s$ ,  $\{y(i - k, j - l) + y(i + k, j + l)\}$  和  $\{g_{k, l}\}$  分别对应写出矢量形式  $\mathbf{z}_s$  和  $\mathbf{g}$ , 则式 (2-176) 可以写为

$$\mathbf{y}_s = \mathbf{g}^T \mathbf{z}_s + \mathbf{e}_s \quad (2-180)$$

则 MRF 模型参数的最小二乘估计为

$$\hat{\mathbf{g}} = \left( \sum_{s \in G} \mathbf{z}_s \mathbf{z}_s^T \right)^{-1} \left( \sum_{s \in G} \mathbf{y}_s \mathbf{z}_s \right) \quad (2-181)$$

$$\hat{\sigma}_e^2 = \frac{1}{n_G} \sum_{s \in G} (\mathbf{y}_s - \mathbf{z}_s^T \hat{\mathbf{g}})^2$$

其中,  $G$  为数据窗;  $n_G$  为该区域中像素的个数。

MRF 模型能很好地逼近纹理, 比较充分地反映出纹理图像局部上下文信息, 以它为基础衍生出了许多纹理表达合成的算法。在实际应用中, MRF 模型对于描述在微观结构上一致的图像纹理具有较好的效果, 诸如草地、沙地或动物的羽毛等, 但对点状、条状或大尺度结构的纹理, 则需要额外的辅助方法加以描述。例如, 为弥补 MRF 模型的不足, 可将小波变换与之结合起来, 建立图像特征的多尺度随机场 (MSRF) 模型。多尺度下的 MRF 模型要比固定尺度下的 MRF 模型具有许多优越性。多尺度下的

MRF 中的马尔可夫链结构是分级构成的, 在模型中无须人为地赋予每个像素点在空间域中的阶数, 较容易实现参数的估计, 解决了 MRF 模型中的归一化常数难题, 另外, 也可以利用参数同时控制粗、精尺度下的图像行为。因此, 多尺度下的 MRF 模型能够更好地精确描述图像行为。

### 2.3.3 局部二值模式

#### 1. LBP 概述

##### 1) LBP 的基本原理

LBP 最初由 Ojala 等<sup>[129]</sup>提出, 该描述符通过刻画图像中每个像素点与其邻域内其他各点的灰度值的差异来描述图像纹理的局部结构特征, 并用一个二进制的数字来量化。这种以邻域为单位的某种局部结构可以看作一个纹理单元, 该纹理单元在整幅图像中有规律地出现就构成了一定的纹理, 而对整幅图像中纹理单元的统计则可以表示整幅图像的纹理特征。从某种意义上说, 该方法将局部的纹理结构信息及全局的纹理统计信息同时融合到纹理分析中, 为同时分析图像中随机的微观纹理和确定的宏观纹理提供了一个有效的工具。

LBP 的基本原理是对于一个  $3 \times 3$  的窗口, 将中心像素与其邻域像素进行比较, 若周围像素值大于中心像素值, 则将该点赋值为 1, 否则赋值为 0, 最后将一个权值模板与阈值处理后的图像进行对应相乘求和, 得到中心像素的值。中心像素的 LBP 值计算方法如图 2.34 所示。LBP 值的计算方法可用下式表示。

$$LBP_{x,y} = \sum_{i=0}^7 s(p_i - p_c) \times 2^i \quad (2-182)$$

其中,  $(x, y)$  表示中心像素坐标;  $s(p_i - p_c) = \begin{cases} 1, & p_i - p_c \geq 0 \\ 0, & \text{其他} \end{cases}$ 。

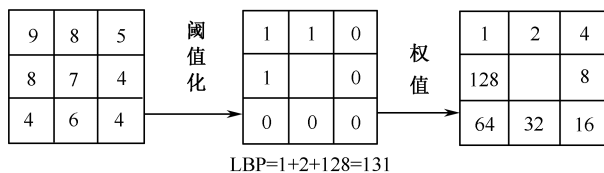


图 2.34 8-邻域 LBP 值的计算方法

统计图像不同 LBP 值出现的概率, 就可以得到图像的纹理谱直方图 (即 LBP 描述符)。显然, 纹理谱直方图为 256 维。

在文献[130]中, Ojala 等将 LBP 进一步扩展至任意圆形邻域  $(P, R)$  (如图 2.35



所示)，其表示如下。

$$\text{LBP}_{P,R} = \sum_{i=0}^{P-1} s(p_i - p_c) \times 2^i \quad (2-183)$$

其中， $P$  表示邻域像素个数； $R$  表示邻域半径。

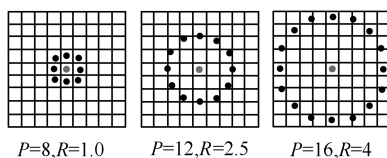


图 2.35 LBP 扩展

## 2) 一致局部二值模式

虽然 LBP 的模式数目很多，但是大多数模式并不能有效地表达局部结构信息，为此 Ojala 等<sup>[130]</sup>引入了一致局部二值模式（Uniform LBP）的概念。目前，很多 LBP 扩展方法是在一致局部二值模式的基础上改进的。一致局部二值模式通过限定 LBP 编码中 0/1 或 1/0 的跳变次数将局部二值模式进行分类。定义在 LBP 中跳变次数不大于 2 的模式为一致模式，其余的所有模式均归为非一致模式。图 2.36 以 8-邻域为例，给出了所有的一致模式。为了与基本局部二值模式区别，采用  $\text{LBP}_{P,R}^{\text{u}2}$  表示一致局部二值模式算子，其中上标“u2”表示 LBP 编码中 0/1 或 1/0 的跳变次数不大于 2。在  $\text{LBP}_{P,R}^{\text{u}2}$  算子中，所有非一致局部二值模式归为一种新的模式，因此该算子共有  $P \cdot (P-1) + 2$  种模式，从而有效地降低了基本局部二值模式的维数。

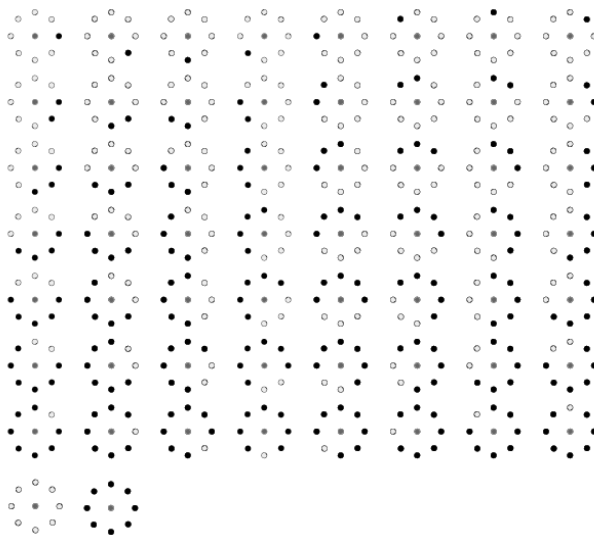


图 2.36 8-邻域一致局部二值模式（黑点表示 1，白点表示 0）

### 3) 旋转不变局部二值模式

图像的旋转变换会导致局部二值模式的二进制编码的循环移位,为此,文献[130]通过将基本局部二值模式编码循环移位到最小,定义了旋转不变局部二值模式,为

$$\text{LBP}_{P,R}^{\text{ri}} = \min \{ \text{ROR}(\text{LBP}_{P,R}, i) | i = 0, 1, 2, \dots, P-1 \} \quad (2-184)$$

其中,  $\text{ROR}(x, i)$  表示将  $P$  位二进制编码的  $x$  循环右移  $i$  位。图 2.37 给出了图像 8-邻域的 36 种旋转不变局部二值模式。

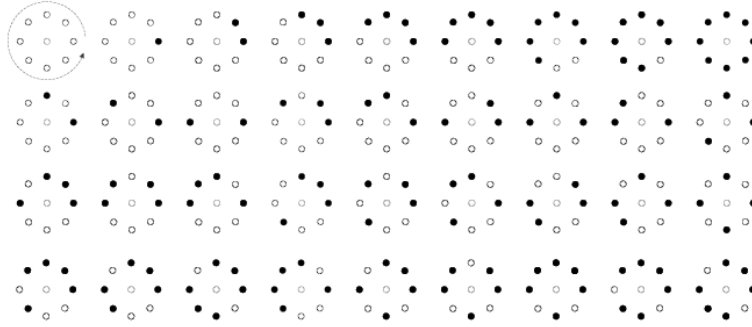


图 2.37 8-邻域旋转不变局部二值模式 (黑点表示 1, 白点表示 0)

### 4) 旋转不变一致局部二值模式

基于一致局部二值模式与旋转不变局部二值模式, Ojala 等<sup>[130]</sup>给出了旋转不变一致局部二值模式的定义, 即

$$\text{LBP}_{P,R}^{\text{riu2}} = \begin{cases} \sum_{i=0}^{P-1} s(p_i - p_c), & U(\text{LBP}_{P,R}) \leq 2 \\ P+1, & \text{其他} \end{cases} \quad (2-185)$$

其中, 上标“riu2”表示旋转不变一致局部二值模式;  $U$  表示在二值模式中由 0 到 1 或由 1 到 0 的转换次数, 其定义为

$$U(\text{LBP}_{P,R}) = |s(p_{P-1} - p_c) - s(p_0 - p_c)| + \sum_{i=1}^{P-1} |s(p_i - p_c) - s(p_{i-1} - p_c)| \quad (2-186)$$

显然,  $\text{LBP}_{P,R}^{\text{riu2}}$  共有  $P+2$  个输出值。图 2.38 给出了 8-邻域基本局部二值模式与旋转不变一致局部二值模式的对应关系。

## 2. LBP 的发展

鉴于其简单有效的特性, 各种针对基本 LBP 扩展的方法层出不穷。目前, 扩展方法主要可以归为以下几类。

(1) 从邻域拓扑结构角度对 LBP 进行扩展。Ojala 等<sup>[130]</sup>将 LBP 扩展到适应任意

的圆形邻域; Liao 等<sup>[131]</sup>将 LBP 扩展到椭圆形邻域, 以满足基于 LBP 描述人脸特征的特殊要求, 同时将圆形邻域作为椭圆形邻域的一种特殊形式; Nanni 等<sup>[132]</sup>研究了用圆、椭圆、抛物线、双曲线、阿基米德螺线等不同的拓扑结构对 LBP 进行扩展; 其他扩展如 MB-LBP (Multi-scale Block LBP)<sup>[133]</sup>, TP-LBP (Three-Patch LBP)、FP-LBP (Four-Patch LBP)<sup>[134]</sup>等。

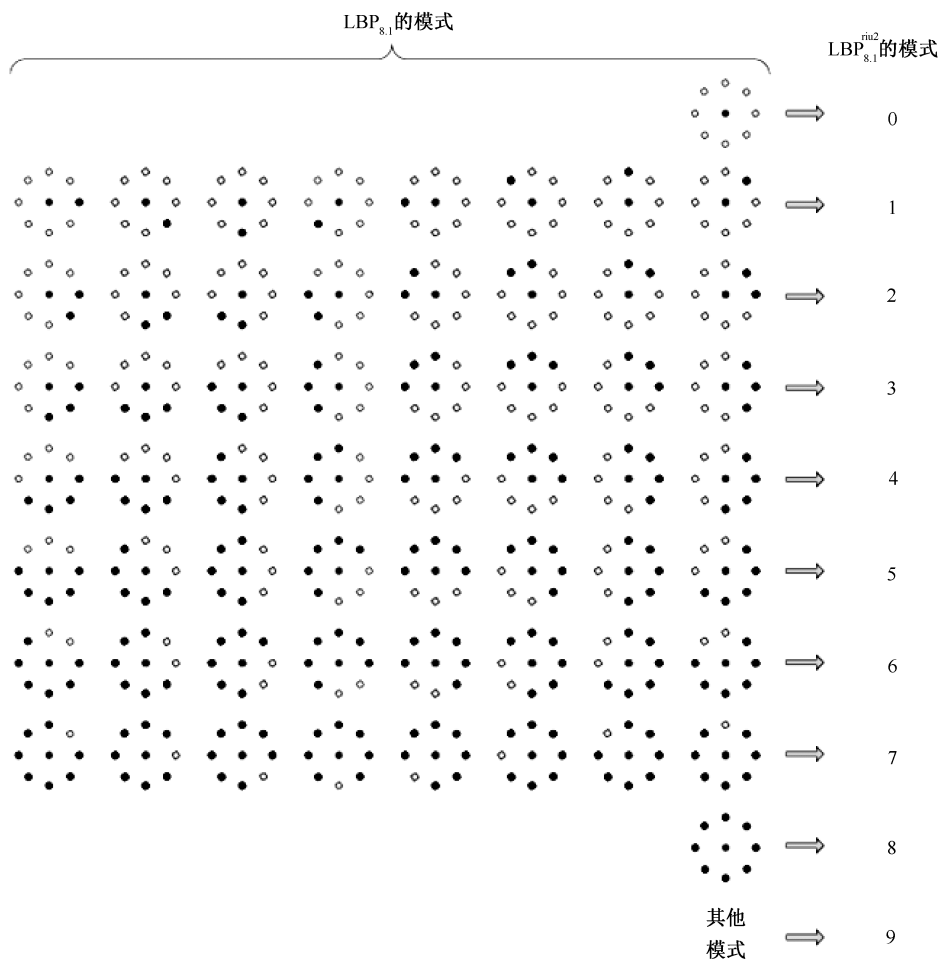


图 2.38 8-邻域 LBP 与  $LBP_{8,1}^{riu2}$  的对应关系 (黑点表示 1, 白点表示 0)

(2) 从降低噪声影响角度对 LBP 进行扩展。Tan 等<sup>[135]</sup>提出了局部三值模式(Local Ternary Pattern, LTP) 来提高基本 LBP 的抗噪能力; Ren 等<sup>[136]</sup>提出了 NRLBP (Noise-Resistant LBP), 以保证在抗噪的同时保持局部邻域的结构特征; Keramidas 等<sup>[137]</sup>提出了模糊局部二值模式来提高 LBP 的抗噪性能; Kylberg 等<sup>[138]</sup>对 8 种 LBP 方法的抗噪性能进行了对比和评价。

(3) 从降维角度对 LBP 进行扩展。Ojala 等<sup>[130]</sup>提出了一致纹理模式 (Uniform LBP) 来降低 LBP 的维数, 将基本 LBP 由 256 维降到 59 维; Heikkilä 等<sup>[139]</sup>提出了中心对称局部二值模式 (Center-Symmetric LBP, CS-LBP) 来降低 LBP 的特征维数; 在 CS-LBP 的基础上, 提出了方向局部二值模式 (Direction LBP, D-LBP)<sup>[140]</sup>, 在实现降维的同时, 进一步提高了 CS-LBP 的分辨能力; Liao 等<sup>[141]</sup>通过选择发生频率高的 LBP 模式 (Dominant LBP) 作为特征, 从而达到降维目的; Zhu 等<sup>[142]</sup>提出了 OC-LBP (Orthogonal Combination LBP), 通过将原邻域划分为多个 4-正交邻域, 并融合每个 4-正交邻域的二值模式值作为最终的描述。

(4) 从编码方式角度对 LBP 进行扩展。Zhao 等<sup>[143]</sup>提出了 CLBC (Completed Local Binary Count) 模式, 该方法采用局部邻域二值化后所包含的 1 的个数作为邻域编码, 进一步降低了 LBP 的计算复杂度; 文献[144]、[145]通过提取局部邻域边缘, 然后基于边缘进行编码; 文献[146]、[147]通过定义局部邻域方向, 提出了基于方向的编码方式; Guo 等<sup>[148]</sup>提出了 CLBP (Completed LBP), 该方法既对局部邻域灰度间的符号变化进行编码, 也对其绝对灰度变化及邻域中心像素进行相应编码; Sapkota 等<sup>[149]</sup>提出了 GRAB (Generalized Region Assigned to Binary) 的编码方式, 首先采用不同分辨率模板对图像进行平滑处理, 然后采用基本 LBP 编码。

(5) 从获取旋转不变性角度对 LBP 进行扩展。Ojala 等<sup>[130]</sup>在将 LBP 扩展到任意圆形邻域、提出一致二值模式的同时, 也提出了旋转不变一致二值模式; Guo 等<sup>[150]</sup>在 LBP 描述符及局部旋转不变量联合分布 (LBP/VAR)<sup>[130]</sup>的基础上, 提出了具有旋转不变性的描述符 LBPV; 在 LBPV 的基础上, 我们对区域局部二值模式也进行了相应的扩展<sup>[151]</sup>。

更多扩展 LBP 的方法可参阅文献[152]、[153]、[154]、[155]、[156]、[157]。

### 3. 几种 LBP 扩展方法

#### 1) LTP

对 LBP 来说, 由于直接采用邻域中心像素的灰度值作为阈值, 所以对噪声比较敏感, 为此 Tan 等<sup>[135]</sup>提出了局部三值模式 (Local Ternary Pattern, LTP), 以减小噪声影响。在 LTP 的定义中, 式 (2-183) 中的函数  $s$  被定义为了如下形式。

$$s(p_i - p_c - \tau) = \begin{cases} 1, & p_i - p_c \geq \tau \\ 0, & |p_i - p_c| < \tau \\ -1, & p_i - p_c < -\tau \end{cases} \quad (2-187)$$

其中,  $\tau$  为阈值。为了降低特征维数, Tan 等又提出将三值模式划分为 Positive LBP

和 Negative LBP 两个二值模式的形式，然后分别计算每一部分的 LBP 值，并综合二者作为邻域特征。LTP 的计算示例如图 2.39 所示。

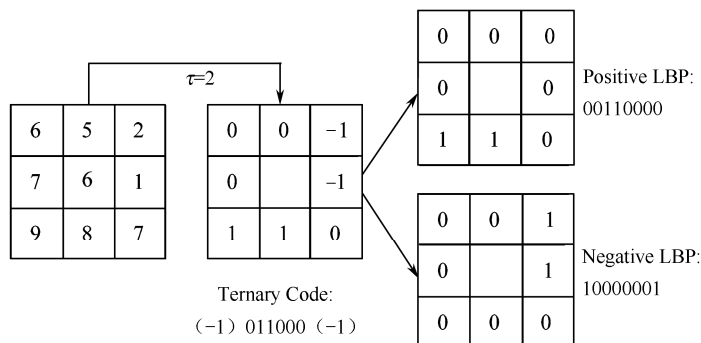


图 2.39 LTP 的计算示例

## 2) CLBP

对 LBP 来说，仅考虑了  $p_i$  与  $p_c$  的大小关系，也即仅考虑了二者灰度差值的符号变化。为了全面描述局部邻域特征，Guo 等<sup>[148]</sup>提出了 CLBP。CLBP 首先基于 LDSMT (Local Difference Sign-Magnitude Transform) 将局部邻域分解为符号和幅值两个部分，即  $p_i - p_c = s_p m_p$ 。其中， $s_p = \text{sign}(p_i - p_c)$ ，代表  $p_i - p_c$  的符号； $m_p = |p_i - p_c|$ ，代表  $p_i - p_c$  的幅值。在此基础上，分别按照 LBP 的方法定义符号变化 ( $s_p$ ) 及幅值变化 ( $m_p$ ) 的二值模式，分别称为 CLBP\_S、CLBP\_M。显然，CLBP\_S 与 LBP 的定义完全一致，因而可以直接用 LBP 表示。幅值变化 ( $m_p$ ) 的二值模式定义为

$$\text{CLBP\_M}_{P,R} = \sum_{i=0}^{P-1} s(m_i - c) \times 2^i \quad (2-188)$$

其中， $c$  为阈值。同时，CLBP 针对基本 LBP 忽略中心像素的问题，定义了 CLBP\_C，即

$$\text{CLBP\_C}_{P,R} = s(p_c - \mu) \quad (2-189)$$

其中， $\mu$  表示图像全局灰度均值。从而 CLBP\_S、CLBP\_M 及 CLBP\_C 构成了 CLBP 描述符。CLBP\_S 及 CLBP\_M 的计算示例如图 2.40 所示。同时，Guo 等人还通过 CLBP\_S、CLBP\_M 及 CLBP\_C 三者相互间的联合分布来提高 CLBP 描述符的分辨能力。

## 3) LBC

不同于传统的二值模式编码方法，Zhao 等<sup>[143]</sup>提出了 LBC 编码方式，LBC 仅仅通过统计二值化后邻域中值为 1 的元素个数作为局部编码，其定义为

$$LBC_{P,R} = \sum_{i=0}^{P-1} s(p_i - p_c) \quad (2-190)$$

结合 CLBP, Zhao 等<sup>[143]</sup>也给出了 CLBC 的定义, 显然 CLBC 也包含 CLBC\_S、CLBC\_M 及 CLBC\_C。其中, CLBC\_S 的定义同 LBC, CLBC\_C 的定义同 CLBP\_C, CLBC\_M 的定义为

$$CLBC\_M_{P,R} = \sum_{i=0}^{P-1} s(m_i - c) \quad (2-191)$$

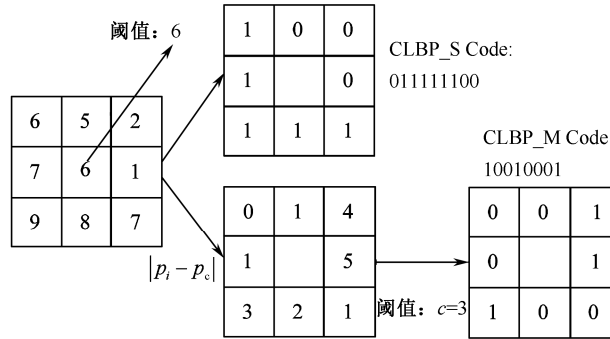


图 2.40 CLBP\_S 及 CLBP\_M 的计算示例

#### 4) CS-LBP

为了有效降低纹理谱描述符的维数, 使其适合用于图像感兴趣区域的描述, Heikkilä 等<sup>[139]</sup>提出了中心对称局部二值模式 (Center-Symmetric Local Binary Pattern, CS-LBP) 描述符, 该方法通过比较 8-邻域与中心像素相对称的 4 对像素间的灰度关系来定义局部纹理模式。使用该方法, 可有效地将传统纹理谱描述符的维数降低到 16 维, 远远低于 LBP 纹理谱描述符的维数; 同时, 该方法还可通过设置全局阈值来判断纹理区域的平坦性。

CS-LBP 值的计算方法为

$$CS-LBP_T(x, y) = \sum_{i=0}^3 s_{CS-LBP}(p_i, p_{i+4}) \times 2^i \quad (2-192)$$

$$s_{CS-LBP}(p_i, p_{i+4}) = \begin{cases} 1, & p_i - p_{i+4} > T \\ 0, & \text{其他} \end{cases}$$

其中,  $T$  为预先设定的阈值, 用于判别局部区域的平坦性。同时, 上述定义也可按照文献[130]的方法扩展到任意的半径及任意邻域像素个数。

#### 5) D-LBP

虽然 CS-LBP 描述符有效地降低了 LBP 纹理谱描述符的维数, 但很明显, 该方法

仅考虑与中心像素对称的 4 对像素间的灰度关系,忽略了中心像素  $p_c$  对局部纹理特征的影响,从而造成了局部纹理信息的丢失;另外,CS-LBP 描述符通过设定全局阈值来判断纹理区域的平坦性,但该阈值很难确定,确定后的阈值也很难适应不同类别的图像。为此,在 CS-LBP 描述符的基础上,我们在文献[140]中提出了方向纹理谱描述符 D-LBP。

为了有效地描述 8-邻域的纹理特征,针对 8-邻域,我们首先给出了以下几个定义。

(1) 如果  $p_i \geq p_c$  ( $i=0,1,2,3$ ), 定义  $p_i \rightarrow p_c$  为正方向, 否则  $p_i \rightarrow p_c$  为反方向。

(2) 如果  $p_c \geq p_{i+4}$  ( $i=0,1,2,3$ ), 定义  $p_c \rightarrow p_{i+4}$  为正方向, 否则  $p_c \rightarrow p_{i+4}$  为反方向。

(3) 如果  $p_i \rightarrow p_c$  及  $p_c \rightarrow p_{i+4}$  同为正方向或同为反方向, 则定义  $p_i$ 、 $p_c$  及  $p_{i+4}$  这 3 个像素同向, 否则 3 个像素为非同向。

(4) 在上述 3 个定义中, 设正方向及同向均用 1 表示, 反方向及非同向均用 0 表示。

按照上述定义, 则方向纹理谱描述符 D-LBP 的数学描述为

$$\text{D-LBP}(x, y) = \sum_{i=0}^3 s_{\text{D-LBP}}(p_i, p_c, p_{i+4}) \times 2^i \quad (2-193)$$

$$s_{\text{D-LBP}}(p_i, p_c, p_{i+4}) = d_1 \cdot d_2 \quad (2-194)$$

$$d_1(p_i, p_c) = \begin{cases} 1, & p_i \geq p_c, i=0,1,2,3 \\ 0, & \text{其他} \end{cases} \quad (2-195)$$

$$d_2(p_{i+4}, p_c) = \begin{cases} 1, & p_{i+4} \leq p_c, i=0,1,2,3 \\ 0, & \text{其他} \end{cases} \quad (2-196)$$

显然, 对于 D-LBP 描述符, 其特征维数仍为 16 维。我们在不提高纹理特征维数的基础上, 进一步考虑了区域中心像素与对称像素间相关性, 因而可以更有效地描述图像的局部纹理特征。同时, 上述定义也可按照文献[130]的方法扩展到任意的半径及任意邻域像素个数。

## 6) 局部边缘二值模式

从式 (2-192) 和式 (2-193) 可以看出, CS-LBP 和 D-LBP 均是通过局部邻域像素间的灰度关系来定义的, 对上述两种局部纹理模式进行进一步扩展, 提出了局部边缘二值模式<sup>[147]</sup>。我们首先给出了局部边缘的定义方法, 与 Sobal、Canny 等边缘检测算子不同, 融合文献[140]关于方向的定义, 这里采用局部邻域像素间的灰度差异来定义局部边缘。针对图像 8-邻域, 局部边缘计算方法如下。

$$\left. \begin{aligned} e_i &= p_i - p_c \\ e_{i+4} &= p_c - p_{i+4} \\ e_c &= p_c - p_c = 0 \end{aligned} \right\} \quad (2-197)$$

其中,  $i \in [0, 3]$ 。当然, 同 LBP 可在不同邻域下扩展一样, CS-LBP、D-LBP 及上述局部边缘的定义都可扩展到不同的邻域下。

结合 CS-LBP、D-LBP 及局部边缘的定义, 我们给出了中心对称局部边缘二值模式 (CS-LEBP) 及方向局部边缘二值模式 (D-LEBP) 的定义。按照 CS-LBP 的定义, CS-LEBP 可表示为

$$\text{CS-LEBP}(x, y) = \sum_{i=0}^3 s'(e_i - e_{i+4}) \times 2^i, \quad s'(t) = \begin{cases} 1, & t \geq 0 \\ 0, & \text{其他} \end{cases} \quad (2-198)$$

结合式 (2-197), CS-LEBP 的定义可进一步表示为

$$\text{CS-LEBP}(x, y) = \sum_{i=0}^3 s'(p_i + p_{i+4} - 2p_c) \times 2^i \quad (2-199)$$

可以看出, 相对于 CS-LBP, CS-LEBP 同时考虑了中心像素的灰度, 因此取得了更好效果。

按照 D-LBP 的定义, D-LEBP 表示为

$$\text{D-LEBP}(x, y) = \sum_{i=0}^3 g(e_i, e_c, e_{i+4}) \times 2^i \quad (2-200)$$

其中,  $g(e_i, e_c, e_{i+4}) = \begin{cases} 1, & (e_i \geq e_c \& e_c \geq e_{i+4}) \parallel (e_i < e_c \& e_c < e_{i+4}) \\ 0, & \text{其他} \end{cases}$ 。结合式 (2-197),

$g(e_i, e_c, e_{i+4})$  进而可以表示为

$$g(e_i, e_c, e_{i+4}) = \begin{cases} 1, & (p_i \geq p_c \& p_{i+4} \geq p_c) \parallel (p_i < p_c \& p_{i+4} < p_c) \\ 0, & \text{其他} \end{cases} \quad (2-201)$$

可以看出, 虽然 CS-LEBP 及 D-LEBP 通过所定义的局部边缘来进行表示, 但仍然体现在局部邻域像素间的灰度变化上, 相比 CS-LBP 及 D-LBP, 计算复杂度及特征维数并没有提高。

## 7) 凹凸局部模式

目前, 在 LBP 的扩展方法中, 关注的重点往往是如何增强从局部邻域所提取特征的分辨能力, 而对空间特征的关注度却不够。图 2.41 给出了一个示例, 其中 (a) 和 (b) 为纹理图像中的两个 8-邻域, 显然 (a) 和 (b) 具有不同的视觉特征, 但是如果按照基本 LBP 及其扩展方法, 它们却得到相同的二值编码, 例如, 具有相同的 LBP 值 219, 具有相同的 LTP 值 (Positive LBP 值为 145, Negative LBP 值为 32), 具有相同的 CLBP\_M 值 177 等。也就是说, 采用这些方法所得到的局部二值模式, 无法区分 (a) 和 (b) 这两个具有完全不同视觉特征的邻域。为此, 我们期望将这些具有不同视觉特征却具有相同二值模式值的邻域区分开来, 因此提出了凹凸局部模式<sup>[148]</sup>。

在纹理分析中, 灰度均值是一种简单有效的纹理统计特征并得到了广泛应用, 因此这里采用邻域灰度均值来描述邻域灰度变化。设  $N(x, y)$  表示一个局部邻域,  $(x, y)$  表



示邻域中心像素的坐标， $\mu_{x,y}$  表示邻域  $N(x,y)$  的灰度均值，其定义为

$$\mu_{x,y} = \frac{1}{P+1} (p_c + \sum_{i=0}^{P-1} p_i) \quad (2-202)$$

显然，对图 2.41 中的邻域 (a) 和 (b) 来说，它们的灰度均值是完全不同的，因此可借助  $\mu_{x,y}$  将它们区分开来。

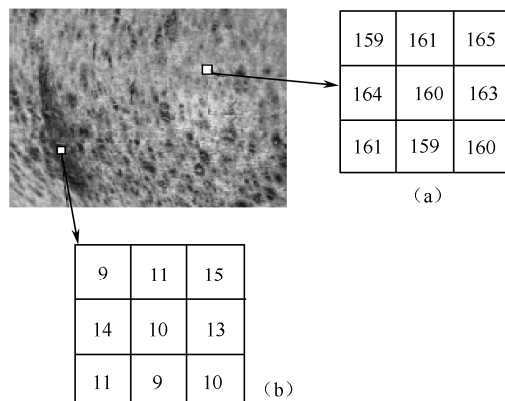


图 2.41 图像 8-邻域示例

借助  $\mu_{x,y}$ ，我们给出了凹凸局部模式的定义：设图像中具有相同二值模式值  $k$  的邻域集合为  $\Omega_k$ ， $\alpha_k$  为阈值，如果邻域  $N(x,y) \in \Omega_k$  且  $\mu_{x,y} < \alpha_k$ ，则定义邻域  $N(x,y)$  为凹 (concave) 邻域，否则为凸 (convex) 邻域，分别统计凹凸邻域的二值模式出现的频率，就得到了新的描述符。

为实现对  $\Omega_k$  的最优凹凸划分，参数  $\alpha_k$  应满足

$$\bar{\alpha}_k = \arg \min_{\alpha_k} \sum_{N(x,y) \in \Omega_k} (\mu_{x,y} - \alpha_k)^2 \quad (2-203)$$

显然，式 (2-203) 满足最小均方误差 (Minimal Mean Square Error, MMSE)，容易得知当  $\bar{\alpha}_k$  满足下式时，划分最优。

$$\bar{\alpha}_k = \frac{1}{|\Omega_k|} \sum_{N(x,y) \in \Omega_k} \mu_{x,y} \quad (2-204)$$

其中， $|\Omega_k|$  表示  $\Omega_k$  中元素的个数。

但如果按照式 (2-204) 的方法，在进行邻域凹凸划分时，对于任二值模式值  $k$ ，都需要重新计算最优参数  $\bar{\alpha}_k$ ，计算复杂度高。对多幅前视红外目标图像统计发现，对于不同二值模式值  $k$ ，其最优参数  $\bar{\alpha}_k$  比较接近。为此，这里可以采用所有最优参数的均值  $\bar{\alpha}$  ( $\bar{\alpha} = \sum_{k=1}^K \bar{\alpha}_k$ ) 代替每一个  $\bar{\alpha}_k$ 。显然，对  $\bar{\alpha}$  来说，也可以直接通过图像所有邻域的灰度均值来计算，即

$$\bar{\alpha} = \frac{1}{\sum_{k=1}^K |\Omega_k|} \sum_{N(x,y) \in \Omega} \mu_{x,y}, \quad \Omega = \bigcup_{k=1}^K \Omega_k \quad (2-205)$$

其中,  $K$  表示图像所具有的不同二值模式值的个数。

设  $\mu$  表示整幅图像的灰度均值, 由式 (2-205) 进一步可知,  $\bar{\alpha} \approx \mu$ , 结合式 (2-204), 显然有  $\bar{\alpha}_k \approx \mu$ 。由此, 我们可以采用  $\mu$  替代  $\bar{\alpha}_k$ , 也就是说, 采用同一阈值  $\mu$  实现对所有二值模式的凹凸划分。实验表明, 采用  $\mu$  代替  $\bar{\alpha}_k$  并不会明显降低凹凸划分的性能, 但却大大降低了计算复杂度。

### 8) OC-LBP

针对 LBP 维数高的问题, 文献[142]提出了 OC-LBP 描述符。OC-LBP 首先将原局部邻域划分为多个 4-正交邻域, 然后采用与 LBP 相同的方法计算每个 4-正交邻域的局部二值模式值, 最后融合所有 4-正交邻域的二值模式值作为最终的描述。设  $M = P/4$  表示局部邻域被划分为 4-正交邻域的个数,  $m = 1, 2, \dots, M$ , OC-LBP 定义为

$$\left. \begin{aligned} \text{OC-LBP1} &= \sum_{i=0}^3 s(p_{M \times i} - p_c) \times 2^i \\ \text{OC-LBP2} &= \sum_{i=0}^3 s(p_{M \times i+1} - p_c) \times 2^i \\ &\dots\dots\dots \\ \text{OC-LBPM} &= \sum_{i=0}^3 s(p_{m+M \times i-1} - p_c) \times 2^i \\ &\dots\dots\dots \\ \text{OC-LBPM} &= \sum_{i=0}^3 s(p_{M+M \times i-1} - p_c) \times 2^i \end{aligned} \right\} \quad (2-206)$$

$$\text{OC-LBP} = [\text{OC-LBP1}, \dots, \text{OC-LBPM}, \dots, \text{OC-LBPM}] \quad (2-207)$$

图 2.42 也给出了针对 8-邻域的 OC-LBP 计算示例。

### 9) CSC-LBP

我们结合纹理的方向特性, 通过局部邻域中心像素与其相对的两个像素间的灰度变化关系定义了该纹理模式, 称为中心对称与中心局部二值模式 (Center-Symmetric and Center Local Binary Pattern, CSC-LBP)。针对图像 8-邻域, 其计算方法为

$$\text{CSC-LBP}(x, y) = \sum_{i=0}^3 s_1(b_i, b_c, b_{i+4}) \times 2^i \quad (2-208)$$

$$s_1(b_i, b_c, b_{i+4}) = b_i \times 2^2 + b_c \times 2^1 + b_{i+4} \times 2^0 \quad (2-209)$$

其中,  $b_i = \begin{cases} 1, & p_i \geq \mu_R \\ 0, & \text{其他} \end{cases}$ ;  $b_c = \begin{cases} 1, & p_c \geq \mu_R \\ 0, & \text{其他} \end{cases}$ 。从式 (2-209) 可知,  $s_1(b_i, b_c, b_{i+4})$  的最

大值为 7, 因此 CSC-LBP 最大可取 105。同时, 由于邻域内各像素的灰度值不可能都小于该邻域的灰度均值, 也即  $b_i$  及  $b_c$  不可能同时取 0, 因而 CSC-LBP 也不可能取 0, 即不存在值为 0 的纹理模式, 也即新纹理谱描述符的维数为 105 维。图 2-43 给出了 CSC-LBP 的计算示例。

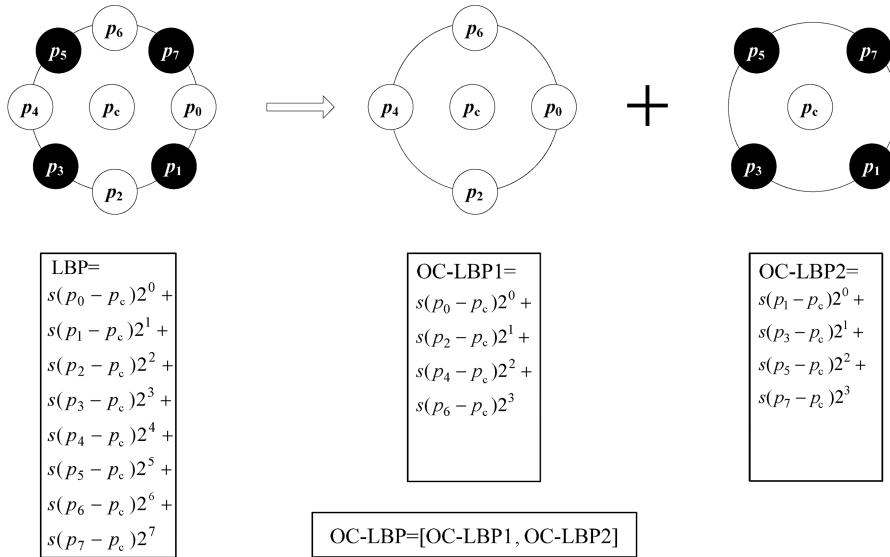


图 2-42 针对 8-邻域的 OC-LBP 计算示例

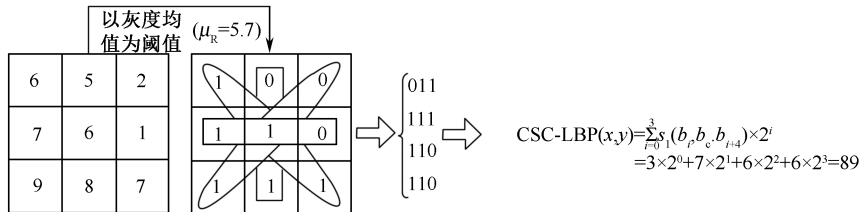


图 2-43 CSC-LBP 的计算示例

## 10) LBPV

在 LBP 方法中局部旋转不变量 (VAR) 是对 LBP 特征的补充。Ojala 等采用 LBP 和局部旋转不变量的联合分布 (LBP/VAR) 作为图像的纹理特征<sup>[130]</sup>。然而, VAR 是一系列连续的值, 因而在使用时要进行量化处理, 并且该处理过程对实验结果影响很大。针对这种情况, 文献[150]提出了 LBPV 描述符, 该方法将图像的空间模式和局部差异性融合为一体, 并克服了 LBP/VAR 存在的问题。LBPV 的定义为

$$\left. \begin{aligned} H(k) &= \sum_{x=1}^N \sum_{y=1}^M s[\text{LBP}(x, y), k], \quad k \in [0, 256] \\ s[\text{LBP}(x, y), k] &= \begin{cases} \text{VAR}(x, y), & \text{LBP}(x, y) = k \\ 0, & \text{其他} \end{cases} \end{aligned} \right\} \quad (2-210)$$

其中,  $k$  表示具体的纹理模式值;  $(x, y)$  表示像素坐标。

由于 VAR 是对区域变化的表示, 所以 VAR 值越大, 对该区域内的区分性贡献就越大, 因而对应该处的编码权重就越大。LBPV 方法无须量化处理, 且完全不需要训练, 因此在纹理分类实验中获得了非常好的结果。但是, LBPV 方法对仿射变化中的不变性问题及对噪声敏感问题并没有给出解决方案。

### 2.3.4 纹理基元共生矩阵

灰度共生矩阵作为一种重要的基于纹理特征进行图像检索的方法, 具有特征提取和相似度计算简便的优点。本小节结合纹理分析方法中的结构分析方法和统计分析方法, 并借鉴方块编码的思想, 构造了一种纹理基元的共生矩阵<sup>[97]</sup>。

#### 1. 方块编码

方块编码 (Block Truncation Coding, BTC), 又称作方块截断编码, 是一种基于块的快速有损图像压缩技术。其基本思想是把一幅图像划分成一系列不重叠的子图像块, 由于子图像块中各像素的灰度相近, 此时只选取两个适当的灰度, 子图像块中的任一像素灰度用这两个灰度中的某一个代替, 于是这个子图像块中的所有像素都映射成这两个灰度值表示的代码, 最后量化成二电平的输出, 这样在重建图像块中就可保持均值和第一个绝对中心矩特性。方块编码算法简单, 性能较高, 信道容错力较好, 计算负担较小, 对存储器要求较少, 重建图像质量较高, 这使其在实时图像传输方面得到很广泛的应用。其编码原理如下。

设定选定的子图像块的两个代表灰度为  $a_0$  和  $a_1$ , 且  $a_1 > a_0$ , 子图像块中第  $i$  个像素的灰度为  $x_i$ , 对于该子图像块的灰度门限  $T$ , 第  $i$  个像素编码后的灰度可按下式计算。

$$Z_i = (1 - y_i)a_0 + y_i a_1 \quad (2-211)$$

$$\text{其中, } y_i = \begin{cases} 1, & x_i \geq T \\ 0, & x_i < T \end{cases}。$$

原子图像块的各像素的灰度  $\{x_1, x_2, \dots, x_m\}$ , 变换后的各像素的灰度  $\{z_1, z_2, \dots, z_m\}$  可以用  $\{a_0, a_1\}$  与  $\{y_1, y_2, \dots, y_m\}$  组合表示, 它们就是子图像的方块编码。借鉴方块编码的思想, 结合人眼的视觉特性, 充分利用图像块中的边缘信息、形状信息及纹理信息来定义图像的纹理基元。针对纹理基元在图像中的分布特征, 结合灰度共生矩阵, 构

造了一种基于纹理基元的共生矩阵方法，用以描述图像的纹理特征。

## 2. 纹理基元的提取

对于彩色图像，在提取纹理基元时，这里采用两种方法首先对彩色图像进行预处理。

(1) 利用式(2-212)转化为灰度图像，然后通过像素之间的灰度差来提取纹理基元。

$$I = 0.3R + 0.59G + 0.11B \quad (2-212)$$

这种方法不需要对彩色图像作过多的处理，简单方便，利用变换后的灰度值来表示原图像的颜色值。

(2) 在某种颜色空间中将颜色进行适当的量化，用量化后的颜色值代替像素的灰度值，通过计算像素间的颜色差来提取图像的纹理基元。这里采用式(2-53)对颜色进行量化。

不同图像的不同区域都有着不同的结构特征，有的区域灰度比较均匀，没有很明显的明暗对比，而有的区域却有着很复杂的灰度差，明暗对比明显。而且，人眼对灰度变化的敏感程度跟背景有关，它随平均灰度的变化而变化，即人眼对图像细节的分辨力与图像的灰度级差有关。当图像本身的灰度级差较小时，人眼的分辨力会降低，反之亦然。因此，可以将图像按照灰度级差分成不同形状的块，来表示图像中的纹理信息。由此，以方块编码的思想为基础，根据图像块的灰度差来进行纹理基元的提取。

纹理基元包含两个要素：其一是该像素区域能表征图像的纹理分布；其二是作为基元，研究的像素数目力求最小。综合这两个要素，算法中规定 $2 \times 2$ 窗口为表征图像纹理的最小单位，称之为纹理基元。具体提取算法如下。

假设 $I$ 代表一幅图像，将 $I$ 划分为 $m \times m$ 大小的互不重叠的子块，对于每个子块，分别计算块内像素的颜色均值 $\mu$ 和平均颜色差 $\sigma$ 。

$$\mu = \frac{\sum_{\forall i,j} I(i,j)}{m^2} \quad (2-213)$$

$$\sigma = \frac{\sum_{\forall i,j} \|I(i,j) - \mu\|}{m^2} \quad (2-214)$$

其中， $I(i,j)$ 表示位于 $(i,j)$ 处的像素的颜色值。

按照方块编码的思想，在每个子块中，对每个像素点，颜色值大于均值 $\mu$ 的赋值为1，反之为0，这样就得到了一系列大小为 $m \times m$ 的二进制块。这些二进制块不仅体现了图像块内的纹理特征，而且在一定程度上反映了图像中的形状分布。相似的纹理结构会产生相同的纹理值，定义这些二进制块为图像的纹理基元。用与这些二进制块等值的十进制值来表示这些纹理基元的值，如图2.44所示，这里取 $m=2$ 。

图像块	<table><tr><td>20</td><td>22</td></tr><tr><td>8</td><td>7</td></tr></table>	20	22	8	7	<table><tr><td>9</td><td>19</td></tr><tr><td>11</td><td>20</td></tr></table>	9	19	11	20	<table><tr><td>17</td><td>11</td></tr><tr><td>18</td><td>9</td></tr></table>	17	11	18	9	<table><tr><td>20</td><td>7</td></tr><tr><td>6</td><td>23</td></tr></table>	20	7	6	23	<table><tr><td>8</td><td>7</td></tr><tr><td>6</td><td>8</td></tr></table>	8	7	6	8
20	22																								
8	7																								
9	19																								
11	20																								
17	11																								
18	9																								
20	7																								
6	23																								
8	7																								
6	8																								
二进制块	<table><tr><td>1</td><td>1</td></tr><tr><td>0</td><td>0</td></tr></table>	1	1	0	0	<table><tr><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td></tr></table>	0	1	0	1	<table><tr><td>1</td><td>0</td></tr><tr><td>1</td><td>0</td></tr></table>	1	0	1	0	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1	<table><tr><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td></tr></table>	1	0	0	1
1	1																								
0	0																								
0	1																								
0	1																								
1	0																								
1	0																								
1	0																								
0	1																								
1	0																								
0	1																								
二进制码	1100	0101	1010	1001	1001																				
纹理基元值	12	5	10	9	9																				
	(a)	(b)	(c)	(d)	(e)																				

图 2.44 图像块与相应的纹理基元值

在提取图像的纹理基元时，会出现如图 2.44 中 (d) 和 (e) 所示的情况，不同结构的块可能会产生相同的纹理基元值。因此，在算法中，需设定一个阈值 $\beta$ ，当图像块的平均灰度差小于这个阈值时，就把这个块看作均匀块，纹理基元值设为 0；大于这个阈值时，就按上述方法计算它的纹理基元值。在提取了图像的纹理基元后，可采用纹理基元值作为整幅图像纹理特征的描述。如图 2.45 所示，(a) 为示例图像，(b) 为用纹理基元表示的示例图像。从图中可以看出，纹理基元在一定程度上可以很好地反映图像的纹理特征。在实际应用中，虽然对图像分块越多越能正确体现图像的特征，但是过细的分割除了造成信息量的急剧增加外，也会导致描述一般性的损失。大量实验证明，将图像划分为 $2 \times 2$ 大小的块是合理的。

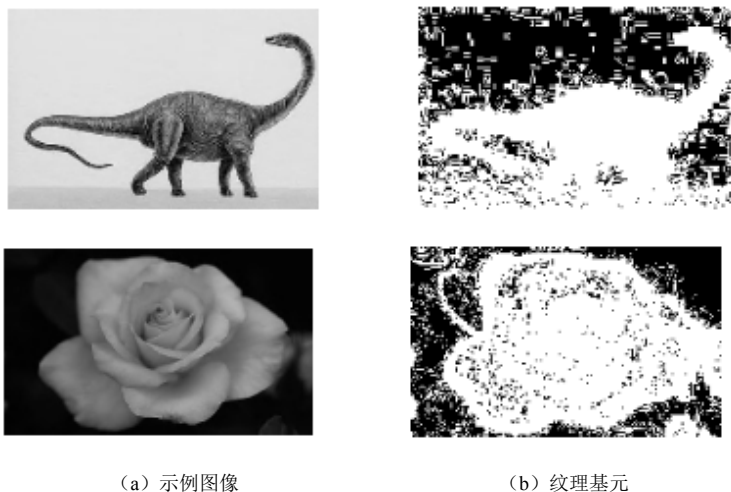


图 2.45 示例图像及其对应的纹理基元

### 3. 构造纹理基元共生矩阵

灰度共生矩阵所提取的图像纹理特征在区分不同的纹理上具有较好的效果, 矩阵中的元素反映了图像灰度分布关于方向、局部邻域和变化幅度的综合信息, 直接处理的基本单元对象是像素, 但它提取的是整幅图像不同方向的纹理特征, 并不能直接提供区别纹理的特征, 还需要从矩阵中进一步提取有意义的统计量。而结构分析法是力求找出纹理基元, 从结构上探索纹理规律, 其基本思想是复杂的纹理可以由一些简单的纹理基元按一定规律重复排列组合而成, 然而这缺乏纹理分布的空间信息。这里利用定义的纹理基元, 结合灰度共生矩阵, 研究其在空间上的分布规律, 构造一种基于纹理基元的共生矩阵, 然后从矩阵中提取有意义的统计量作为图像的纹理特征, 从而将纹理分析中的结构法和统计法有机地结合起来。

设  $I'$  是  $I$  提取了纹理基元大小的图像, 图像中的纹理基元值为  $[0, n-1]$ , 则基于纹理基元的共生矩阵中的元素为

$$H(i, j) = \eta \left[ p(x, y), p(N_{(x, y)}) \right] = \alpha \sum_{x=1}^N \sum_{y=1}^M C_i(x, y) \sum_{(x', y') \in N(x, y)} C_j(x', y') \quad (2-215)$$

其中,  $i, j = 0, 1, 2, \dots, n-1$ ;  $p(x, y)$  是  $I'$  中  $(x, y)$  处的纹理基元值;  $N_{(x, y)}$  表示  $I'$  中  $(x, y)$  的 4-邻域;  $p(N_{(x, y)})$  表示  $I'$  中  $(x, y)$  的 4-邻域处的纹理基元值。公式中  $C_i(x, y)$  的取值如下。

$$C_i(x, y) = \begin{cases} 1, & p(x, y) = i \\ 0, & \text{其他} \end{cases} \quad (2-216)$$

于是, 改进的共生矩阵中的元素  $H(i, j)$  就表示  $I'$  中所有  $(x, y)$  处的像素与其 4-邻域内的像素构成的像素对等于  $(i, j)$  的数目。考虑到图像的大小对矩阵的影响, 对矩阵进行归一化处理。

用该矩阵描述图像的纹理, 避免了传统的共生矩阵只能描述单方向信息而带来的问题。该共生矩阵反映了图像灰度分布关于方向、局部邻域和变化幅度的综合信息, 但它并不能直接提供区别纹理的特征, 还需要从中进一步提取有用的统计量构成纹理特征。为此, 这里要提取 4 个典型的统计量 (能量、对比度、相关性和熵) 来反映纹理不同方面的特征。由上述 4 个特征组成图像的纹理特征向量来分析及比较含有纹理结构的图像。

## 2.4 MPEG-7 中的图像特征描述符

面对如何有效地识别、过滤、浏览和检索视听材料的迫切要求, MPEG 于 1996 年开始从事一项新的工作, 目的是为多媒体内容的描述产生一个标准。MPEG 的这个

新成员被正式命名为“多媒体内容描述接口 (multimedia content description interface)”，简称 MPEG-7。其目标是产生一个描述多媒体内容的标准，支持对多媒体信息在不同程度层面上的解释和理解，从而使其可以根据用户的需要进行传递和存取。MPEG-7 并不面向某种具体的应用，相反，MPEG-7 标准将支持尽可能广泛的应用领域。MPEG-7 标准中确定了一个标准描述子 (descriptors，也称描述符) 集，用来描述各种类型的多媒体信息。描述子是特征的描述，它定义了特征描述的句法和语义。一个特征，如颜色、纹理或形状，可能有多个描述子对其不同相关方面进行描述。如图 2.46 所示，MPEG-7 注重的是提供视听信息内容的描述方案，并不包括针对不同应用的特征提取算法和搜索引擎。这使得 MPEG-7 标准一方面可以被广泛地应用，不局限于某些与特殊应用密切相关的特征提取算法和搜索引擎，也不依赖于被描述内容的编码和存储方式；另一方面又可以引入竞争机制，使人们能够针对不同应用领域产生出更多更好的特征提取算法和搜索引擎。有关 MPEG-7 的更详细介绍可参阅 <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm> 及 ISO/IEC 颁发的有关 MPEG-7 的文档。

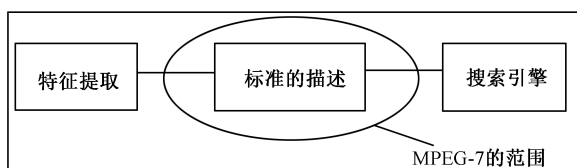


图 2.46 MPEG-7 的范围

MPEG-7 标准包含几个重要概念：描述符 (Descriptors, D)、描述方案 (Description Schemes, DS) 及描述定义语言 (Description Definition Language, DDL)。其中，描述符表示多媒体特征信息的语法和语义属性，一个特征可以用多个描述符来表示。例如，平均色 (average color)、主色 (dominant color) 和颜色直方图 (color histogram) 等都是颜色特征的描述符。描述方案指定了对象或者特征的结构和关系，一般情况下，描述方案是解决图像分类和组织问题或以特定的索引结构描述图像内容的子系统。描述定义语言允许创建新的描述方案和描述符，也允许扩展和修改现存的描述方案。目前，MPEG-7 采用带有 MPEG-7 特定扩展的 XML (可扩展置标语言) 语言作为描述定义语言。

在 MPEG-7 标准中考虑了 5 类基本的视觉特征，对应地使用了 5 类描述符：颜色描述符、形状描述符、纹理描述符、运动描述符和位置描述符。其中，前 3 类描述符都是针对单幅静止图像的描述。由于 MPEG-7 的目的是建立对图像信息完备而又一致的描述，因此，MPEG-7 中提出的颜色描述符、纹理描述符和形状描述符等单类描述符必须尽可能地完备而又一致，也就是说，使用这些描述符能尽可能完备地表达图像中的视觉信息，并且更重要的是这些描述符适用于所有图像。



下面主要介绍一下 MPEG-7 中有关图像内容描述的 3 类描述符：颜色描述符、形状描述符及纹理描述符<sup>[1]</sup>。

### 2.4.1 颜色描述符

MPEG-7 中的颜色描述符有颜色空间描述符（color space descriptor）、颜色量化描述符（color quantization descriptor）、主颜色描述符（dominant color descriptor）、可伸缩颜色描述符（scalable color descriptor）、颜色布局描述符（color layout descriptor）、颜色结构描述符（color-structure descriptor）及帧图/图组颜色描述符（group of frames/group of pictures color descriptor）等。其中，颜色空间描述符和颜色量化描述符是两个辅助性的颜色描述符，它们不能独立使用，只能配合其他颜色描述符使用。

颜色空间描述符描述了 MPEG-7 颜色描述符的颜色空间，包括 RGB、YCrCb、HMMD、HSV 及各种颜色系统与 RGB 的线性变换矩阵。

颜色量化描述符描述了颜色空间的均匀量化，量化产生的维（Bin）的数目是可配置的，这样使得各种应用具有更大的灵活性。该描述符往往需要和主颜色描述符等配合使用。

主颜色描述符最适用于表示局部（对象或图像区域）特征，几种颜色就足以表达我们感兴趣区域的颜色信息。当然，它也可以用于整个图像，如旗帜图像或彩色商标图像。颜色量化用于提取每个区域/图像的少数代表颜色，并相应地计算出区域中每种量化颜色所占的百分比。同时，该描述符还定义了整个描述符的空间相关性，用于相似性检索。

可伸缩颜色描述符定义了 HSV 空间的颜色直方图，然后用 Harr 变换编码。根据 Bin 的数目和 Bit 表示的精度，它的二进制表达在 Bin 的数量和 Bit 的表达精度上都是可伸缩的。这个描述符主要用于图像与图像的匹配和基于颜色特征的检索，检索的精度随着描述中使用的比特数目的增加而提高。

颜色布局描述符描述了整幅图像或者图像的部分区域的颜色空间分布状况。该描述符以一种紧凑的形式，有效地表达了颜色的空间分布。这种紧凑性以很小的计算代价，带来高速的浏览和检索。它提供图像与图像的匹配和超高速的片断与片断的匹配，这些匹配要求大量相似性计算的重复。由于该描述符表达了颜色特征的布局信息，因此它可以提供相当友好的用户接口，如使用其他颜色描述符中均不支持的手绘草图查询。

颜色结构描述符是一个颜色特征描述符，它既包括颜色内容信息（类似于颜色直方图），又包括内容的结构信息。它的主要功能是图像与图像的匹配，主要用于静态图像检索。在这里一幅图像可能由一个单一-矩形或者任意形状，也可能是非连通的区

域组成。提取方法是通过考虑一个  $8 \times 8$  像素的结构化元素中的所有颜色，将颜色结构信息加入该描述符中，而不是单独考虑每个像素。

帧图/图组颜色描述符用于静态图像的可伸缩颜色描述符扩展到对视频片断或静态图像集合的颜色描述。在 Haar 变换之前，用附加的两个比特定义如何计算颜色直方图，是均值、中值还是相交。

## 2.4.2 形状描述符

MPEG-7 中定义的形状描述符有区域形状描述符 (region shape descriptor)、轮廓形状描述符 (contour shape descriptor) 及三维形状描述符 (shape 3D) 3 种。

物体的形状可能是一个独立的整体，也可能是由几个区域构成的。由于使用了图像中构成物体形状的所有像素信息来描述物体，区域形状描述符不仅可以用来刻画出个别独立的连通区域，也包括由一些不连通区域所构成的复杂图形，如图 2.47 所示。

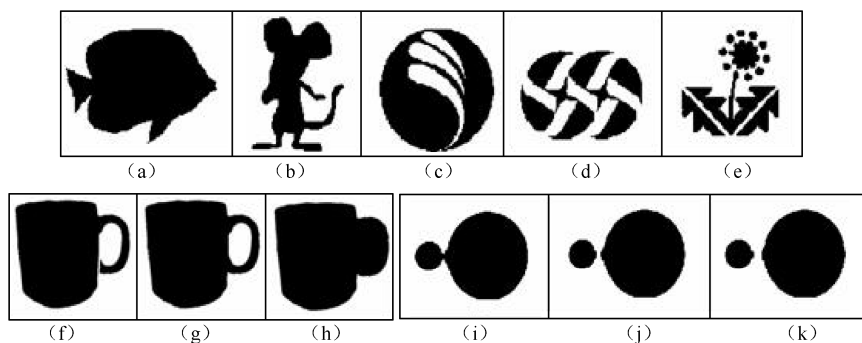


图 2.47 各类形状示例图

区域形状描述符的表达式是由一系列 ART 系数构成的，ART 定义了一组二维的复值正交基函数，将二维区域投射到这些基函数上，得到的系数归一化后就可以描述区域的形状并用于匹配。它也是一种非常紧凑、有效的描述方式，并具备分割噪声的功能。

轮廓形状描述符是利用轮廓的曲率尺度空间来描述封闭的轮廓。具体来说，对封闭轮廓线进行多尺度变换，得到多个尺度的轮廓线，随着尺度的变换，轮廓线越来越平滑。多尺度变换可以使用高斯函数对曲线上的点进行平滑来完成。然后，在每一个尺度计算曲线的曲率，得到曲率变化的零交叉点，使用这些点就可以描述轮廓的形状。

MPEG-7 中的三维形状描述符可用于相对自然的或虚拟的三维目标。在描述三维物体的形状特征时，首先建立物体的三维网格 (3D mesh) 模型，然后在物体表面的局部区域计算出该处的形状指数 (shape index)。该描述符基于形状频谱 (shape

spectrum) 的概念, 三维形状描述符事实上就是基于形状指数的直方图。三维形状描述符常常用于三维物体的匹配。

### 2.4.3 纹理描述符

MPEG-7 中的纹理描述符包括同质纹理描述符 (homogenous texture descriptor)、纹理浏览描述符 (texture browsing descriptor) 和边缘直方图描述符 (edge histogram descriptor) 3 种。同质纹理描述符在纹理具有一致性的区域统计纹理的空间频率。纹理浏览描述符从纹理的方向性、规则性和粗糙程度 3 个方面进行描述。边缘直方图描述符在纹理不具有一致性的区域描述了边缘的空间分布。

同质纹理描述符通过在频域计算能量和能量方差来提供对纹理的量化描述。同质纹理描述符采用 5 个尺度和 6 个方向的 30 个 Gabor 滤波器对纹理图像进行多分辨率分解, 将频域内滤波器组输出能量的均值和标准差作为纹理特征。它描述了图像中与人类感知一致的规则性、方向性和粗糙程度等纹理特征, 最适合对具有同质特征纹理进行一定的描述, 可用于纹理图像数据库中图像之间的相似性匹配, 因此主要用于大量相似图案的搜索和浏览。一幅图像可看作由同质纹理以马赛克形式拼接而成, 所以与这些区域关联的纹理特征可以作为索引来检索图像。例如, 用户浏览一个航空图像数据库, 可能想识别图像集合中的停车场, 当从远处观察时, 汽车规则 (以相等间隔) 停放的停车场就是一个极好的同质纹理图案的例子。同样, 从空中拍摄或是卫星拍摄的农田和植被图像也是同质纹理的示例。该描述符可以支持诸如“找出一个看起来与这个区域相似的植被”的查询。为了支持这样的图像检索, 需要对纹理进行有效的表示。

纹理浏览描述符从类似于人类感知的角度对纹理的方向性、规则性和粗糙程度进行描述, 适用于图像的浏览和根据纹理粗糙程度进行的分类。由于一个纹理可能不只包含一个主要方向和相应的尺度, 因此允许最多指定两个不同的方向和粗糙度值。该描述符的计算方法和同质纹理描述符类似, 首先使用一组带有方向和尺度参数的 Gabor 滤波器进行滤波, 然后通过分析滤波结果, 找到纹理主要的方向, 接着分析滤波后的图像沿着这两个 (第二个主方向是可选的) 主方向投影, 以此决定纹理的方向性、粗糙度和规则性。同质纹理描述符和纹理浏览描述符提供了表示相似纹理区域的多尺度方法。这是一个非常紧凑的描述符, 适用于浏览应用和纹理过滤, 与同质纹理描述符联合使用可以实现快速准确的图像检索。

边缘直方图描述符描述边缘的空间分布信息, 首先将图像划分成 16 个互不重叠的矩形区域, 对每个图像区域分别按水平、垂直、 $45^\circ$  角、 $135^\circ$  角 4 个方向和 1 个无方向性

边缘 5 类信息（如图 2.48 所示）进行直方图统计。此描述符具有尺度不变性，支持纹理旋转和旋转不变匹配，适用于非一致纹理图像，如普通图像的检索。

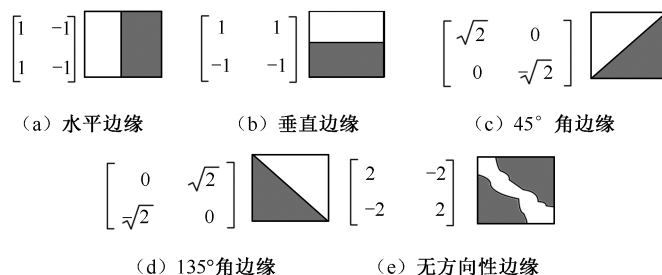


图 2.48 边缘类型定义

## 参 考 文 献

- [1] 孙君顶, 赵珊. 图像低层特征提取与检索技术[M]. 北京: 电子工业出版社, 2009.
- [2] 孙兴华. 基于内容的图像检索研究[D]. 南京: 南京理工大学, 2001.
- [3] 黄元元. 基于视觉特征的图像检索技术研究[D]. 南京: 南京理工大学, 2003.
- [4] Oleg V. Color Image quantization in windows systems with local K-means algorithm[C] // Proceedings of VI Western Computer Graphics Symposium, 1995:74-79.
- [5] Zhou B, Shen J Y, Peng Q B. An adjustable algorithm for color quantization[J]. Pattern Recognition Letters, 2004, 25(16):1787-1797.
- [6] Scheunders P. A genetic approach towards optimal color image quantization[J]. Image Processing, 1996, 7(5):1031-1034
- [7] Gerrautz M, Purgathofer W. A simple method for color quantization: octree quantization [C] // Proceedings of ICG'98, 1998, 8(6):219-230.
- [8] Celebi M E, Wen Q, Hwang S, et al. Color Quantization of Dermoscopy Images Using the K-Means Clustering Algorithm[C] // Color Medical Image Analysis. Springer Netherlands, 2013:87-107.
- [9] Wen Q, Celebi M E. Hard versus fuzzy c-means clustering for color quantization[J]. EURASIP Journal on Advances in Signal Processing, 2011(1):1-12.
- [10] Androustos D, Plataniotis K N, Venetsanopoulos A N. Image retrieval using the directional detail histogram[J]. SPIE 3312, 1997:129-137.
- [11] 王涛, 胡事民, 孙家广. 基于颜色-空间的图像检索[J]. 软件学报, 2002, 13(10): 2031-2036.

- [12] 孙君顶, 毋小省. 基于分块主色和形状特征的彩色图像检索[J]. 光电工程, 2006, 33(12):85-90.
- [13] M. Emre Celebi. Improving the performance of k-means for color quantization[J]. Image and Vision Computing, 2011, 29(4):260-271.
- [14] Celebi M E, Wen Q. VARIANCE-CUT: A fast color quantization method based on hierarchical clustering[J]. ICECCO, IEEE, 2013:103-106.
- [15] Palomo E J, Domínguez E. Hierarchical Color Quantization Based on Self-organization[J]. Journal of Mathematical Imaging and Vision, 2013:1-19.
- [16] Yue X D, Miao D Q, Cao L B, et al. An efficient color quantization based on generic roughness measure[J]. Pattern Recognition, 2014, 47(4):1777-1789.
- [17] Schaefer G. Soft computing-based colour quantisation[J]. EURASIP Journal on Image and Video Processing, 2014, 2014(1):8.
- [18] Stricker M, Orengo M. Similarity of color images[C] // Proceedings of SPIE Storage and Retrieval for Image and Video Database, 1995, 2420:381-392.
- [19] 刘忠伟. 利用局部累加直方图进行彩色图像检索[J]. 中国图像图形学报, 1998, 3(7):533-537.
- [20] 梁艳梅, 翟宏坤, 毋国光. 基于模糊相关的彩色图像检索[J]. 中国科学(E辑), 2003, 33(10):934-938.
- [21] 王炜, Michael R L, 武德峰. 基于模糊分类的图像颜色直方图研究[J]. 模糊系统与数学, 2003, 17(4):94-98.
- [22] Chai L, Qin Z, Zhang H, et al. A new multi-scale fuzzy model for Histogram-Based Descriptors[J]. ICMEW, 2013, 2013:1-6.
- [23] Liu G H, Yang J Y. Content-based image retrieval using color difference histogram[J]. Pattern Recognition, 2013, 46(1):188-198.
- [24] Min R, Cheng H D. Effective image retrieval using dominant color descriptor and fuzzy support vector machine[J]. Pattern Recognition, 2009, 42(1):147-157.
- [25] 孙君顶, 毋小省. 基于分块主色和形状特征的彩色图像检索[J]. 光电工程, 2006, 33(12):85-90.
- [26] Talib A, Mahmuddin M, Husni H, et al. A weighted dominant color descriptor for content-based image retrieval[J]. Journal of Visual Communication and Image Representation, 2013, 24(3):345-360.
- [27] 孙君顶. 基于内容的图像检索技术[D]. 西安: 西安电子科技大学, 2005.
- [28] 赵珊. 基于内容的图像检索关键技术研究[D]. 西安: 西安电子科技大学, 2007.
- [29] Funt B V, Finlayson G D. Color constant color indexing[J]. IEEE Transactions on Pattern Analysis and Maching Intelligence, 1995, 17:522-529.
- [30] Gevers T, Stokman H M G. Robust histogram construction from color invariants for

- object recognition[J]. IEEE PAMI, 2004, 26(1):113-118.
- [31] Gijssenij A, Gevers T, Van De Weijer J. Computational color constancy: Survey and experiments[J]. IEEE Transactions on Image Processing, 2011, 20(9):2475-2489.
- [32] John Z M. An Information theoretic approach to content based image retrieval[D]. Baton Rouge: Louisiana State University and Agricultural and Mechanical College, 2000.
- [33] Constantium N, Nozha B. Spatially constrained color distributions for image indexing[J]. CGIP, 2000:261-265.
- [34] Qiu G P, Lam K M. Frequency layered color indexing for content-based image retrieval [J]. IEEE Transactions on Image Processing, 2003, 12(1):102-113.
- [35] 宋擒豹, 杨向荣, 沈钧毅, 等. 图像相似模式挖掘中的颜色-位置直方图方法[J]. 计算机研究与发展, 2002, 39(9):1132-1137.
- [36] Pass G, Zabih R, Miller J. Comparing images using color coherence vectors[C] // Proceedings of the fourth ACM international conference on Multimedia, 1997:65-73.
- [37] Huang J. Color-spatial image indexing and applications[D]. New York Cornell University, 1998.
- [38] Hsu W, Chua T S, Pung H K. An integrated color-spatial approach to content-based image retrieval[C] // Proceedings of ACM Multimedia'95 Conference, San Francisco, 1995:305-313.
- [39] Yoo H W, Jang D S, NA Y K. An efficient indexing structure and image representation for content-based image retrieval[J]. IEICE Trans. on INF&SYST, 2002:1390-1398.
- [40] Stehling R O, Nascimento M A, Falcao A X. On 'shapes' of colors for content-based image retrieval[C] // The ACM Multimedia Conference in Los Angeles, 2002:171-174.
- [41] 何清法, 李国杰. 综合分块主色和相关反馈技术的图像检索方法[J]. 计算机辅助设计与图形学学报, 2001, 13(10):912-917.
- [42] Rao A, Srihari R, Zhang Z. Spatial Color Histogram for Content-Based Retrieval[C] // In 11th IEEE International Conference on Tools with AI, 1999:183-186.
- [43] Lim S, Lu G J. Spatial statistics for content-based image retrieval[C] // Proceedings of the International Conference on Information Technology, 2013.
- [44] 张旭, 郭宝龙, 孟繁杰, 等. 基于 IPDSH 兴趣点空间区域划分的图像检索[J]. 吉林大学学报:工学版, 2013 (5):1408-1414.
- [45] 孟繁杰, 郭宝龙, 李新伟, 等. 基于兴趣点凸包的图像检索方法[J]. 光电子·激光, 2010, 21(6):936-939.
- [46] Sun J D, Zhang X M, Cui J T, et al. Image retrieval based on color distribution entropy[J]. Pattern Recognition Letters, 2006, 27(10):1122-1126.

- [47] Sun J D. Image Retrieval Based on Improved Entropy and Moments[C] // International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2006.
- [48] 孙君顶, 毋小省. 基于位平面熵及分布熵的图像检索[J]. 系统工程与电子技术, 2009, 31(3):719-722.
- [49] Yong D C, Sang Y S, Nam C K. Image retrieval using BDIP and BVLC moments[J]. IEEE transactions and systems for video technology, 2003, 13(9):951-957.
- [50] Ritendra D, Dhiraj J, Li J, et al. Image retrieval: ideas, influences, and trends of the new age[J]. ACM Transactions on Computing Survey, 2008, 40(2):1-66.
- [51] Zhang D S, Lu G J. Review of shape representation and description techniques[J]. Pattern Recognition, 2004, 37:1-19.
- [52] Amanatiadis A, Kaburlasos V G, Gasteratos A, et al. Evaluation of shape descriptors for shape-based image retrieval[J]. IET: Image Processing, 2011, 5(5):493-499.
- [53] 周明全, 耿国华, 韦娜. 基于内容图像检索技术[M]. 北京: 清华大学出版社, 2007.
- [54] Veltkamp R C, Latecki L J. Properties and performance of shape similarity measures [M]. Springer Berlin Heidelberg, 2006.
- [55] 庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索[M]. 北京: 清华大学出版社, 2002.
- [56] Mokhtarian F, Bober M. Curvature scale space representation: theory, applications, and MPEG-7 standardization[M]. Springer Publishing Company, Incorporated, 2011.
- [57] Kim W H, Pachauri D, Hatt C, et al. Wavelet based multi-scale shape features on arbitrary surfaces for cortical thickness discrimination[C] // NIPS, 2012:1250-1258.
- [58] 杨翔英, 章毓晋. 小波轮廓描述符及在图像查询中的应用[J]. 计算机学报, 1999, 22(7):752-757.
- [59] Freeman H. Comparative analysis of line drawing modeling schemes[J]. Computer Graphics Image Process, 1980, 12:203-223.
- [60] Liu Y K, Zalik B. An efficient chain code with Human coding[J]. Pattern Recognition, 2005, 38:669-676.
- [61] Li J, Guo S, Ye F. Shape recognition based on Freeman chain code[J]. Advanced Materials Research, 2011, 317:2490-2496.
- [62] Lemus E, Bribiesca E, Garduño E. Representation of enclosing surfaces from simple voxelized objects by means of a chain code[J]. Pattern Recognition, 2014, 47(4): 1721-1730.
- [63] 赵宇, 陈雁秋. 曲线描述的一种方法: 夹角链码[J]. 软件学报, 2004, 15(4):300-307.
- [64] 刘淑娟. 可变夹角链码的研究[D]. 石家庄: 河北师范大学, 2005.

- [65] Hermilo S C, Ernesto B, Ramon R D. Efficiency of chain codes to represent binary objects[J]. Pattern Recognition, 2007, 40:1660-1674.
- [66] Sánchez-Cruz H, López-Valdez H H. Equivalence of chain codes[J]. Journal of Electronic Imaging, 2014, 23(1).
- [67] Iivarinen J, Visa A. Shape recognition of irregular objects[C] // Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, SPIE, 1996:25-32.
- [68] 王小玲, 谢康林. 一种新的方向码描述的图像检索方法[J]. 哈尔滨工业大学学报, 2006, 38(9):1545-1548.
- [69] Sun J D, Wu X S. Shape retrieval based on the relativity of chain codes[J]. MCAM' 2007:76-84.
- [70] 孙君顶. 基于链码分布特征及相关性的轮廓描述与检索[J]. 光电子·激光, 2008, 19(8):1112-1115.
- [71] Wu X S, Sun J D. Shape retrieval of irregular objects[J]. ITESS, 2008, 3:319-322.
- [72] He X C, Yung N H. Curvature scale space corner detector with adaptive threshold and dynamic region of support[C] // 17th IEEE International Conference on Pattern Recognition, 2004:791-794.
- [73] 张小洪, 雷明, 杨丹. 基于多尺度曲率乘积的鲁棒图像角点检测[J]. 中国图像图形学报, 2007, 7(12):1270-1275.
- [74] 张兆生. 形状特征提取及检索技术研究[D]. 河南: 河南理工大学, 2008.
- [75] Chen C C. Improved moment invariants for shape discrimination[J]. Pattern Recognition, 1993, 26(5):683-686.
- [76] Gupta L, Srinath, M D. Contour sequence moments for the classification of closed planar shapes[J]. Pattern Recognition, 1987, 20(3):267-272.
- [77] 曹茂永, 孙衣亮, 郁道银. 用于模式识别的极半径不变矩[J]. 计算机学报, 2004, 6(27):860-864.
- [78] Hu M K. Visual pattern recognition by moment invariants[J]. IRE Transactions on Information Theory, 1962, 8(2):179-187.
- [79] Papakostas G A, Karakasis E G, Koulouriotis D E. Novel moment invariants for improved classification performance in computer vision applications[J]. Pattern Recognition, 2010, 43(1):58-68.
- [80] Flusser J, Kautsky J, Šroubek F. Implicit moment invariants[J]. International Journal of Computer Vision, 2010, 86(1):72-86.
- [81] 刘进, 张天序. 图像不变矩的推广[J]. 计算机学报, 2004, 27(5):668-674.
- [82] Teague M R. Image analysis via the general theory of moments[J]. J. Opt. Soc. Am. 1980, 70 (8):920-930.



- [83] Chen Z, Sun S K. A Zernike moment phase-based descriptor for local image representation and matching[J]. IEEE Transactions on Image Processing, 2010, 19(1): 205-219.
- [84] Singh C, Mittal N, Walia E. Face recognition using Zernike and complex Zernike moment features[J]. Pattern Recognition and Image Analysis, 2011, 21(1):71-81.
- [85] Bober M. Mpeg-7 visual shape descriptor[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(6):716-719.
- [86] Zhang D S. Image Retrieval Based on Shape[D]. Metbourne: Monash University, 2002.
- [87] 孙君顶, 崔江涛, 毋小省, 等. 基于颜色和形状特征的彩色图像检索方法[J]. 中国图像图形学报, 2004, 9(7):820-827.
- [88] 林元烈. 应用随机过程[M]. 北京: 清华大学出版社, 2002:78-130.
- [89] 孙君顶, 武学东, 周利华. 基于颜色和形状的图像检索[J]. 计算机科学, 2004, 31(5):180-183.
- [90] Lu G J, Sajjanhar. A Region-based Shape Representation and Similarity Measure Suitable for Content-based Image Retrieval[J]. Multimedia System, 1999, 7(2):165-174.
- [91] 孙君顶, 毋小省. 基于熵及不变矩特征的图像检索[J]. 光电工程, 2007, 34(6): 102-106+115.
- [92] Coggins J M, Jain A K. A spatial filtering approach to texture analysis[J]. Pattern Recognition, 1985, 3:195-203.
- [93] Castleman K R. 数字图像处理[M]. 朱志刚, 等译. 北京: 电子工业出版社, 2002.
- [94] 马媛媛. 基于纹理分类的图像检索技术研究[D]. 河南: 河南理工大学, 2011.
- [95] Haralick R M, Shanmugam K. Texture features for image classification[J]. IEEE Transactions on System, Man and Cybernetics, 1973, 3(6):610-621.
- [96] 洪继光. 灰度-梯度共生矩阵纹理分析方法[J]. 自动化学报, 1984, 10(1):22-25.
- [97] 赵珊, 孙君顶, 周利华. 基于方块编码的图像纹理特征提取及检索算法[J]. 光电子·激光, 2006, 17(8):1014-1017.
- [98] Roberti de Siqueira F, Robson Schwartz W, Pedrini H. Multi-scale gray level co-occurrence matrices for texture description[J]. Neurocomputing, 2013, 120:336-345.
- [99] Liu G, Wang R, Deng Y K, et al. A New Quality Map for 2-D Phase Unwrapping Based on Gray Level Co-Occurrence Matrix[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(2):444-448.
- [100] Tamura H, Mori S, Yamawaki T. Texture feature corresponding to visual perception[J]. IEEE-SMC, 1978, 8(6):460-473.
- [101] 盛文, 杨江平, 柳建, 等. 一种基于纹理元灰度模式统计的图像纹理分析方法

- [J]. 电子学报,2000,28(4).
- [102] Xie X, Mirmehdi M. TEXEM: Texture exemplars for defect detection on random textured surfaces[J]. IEEE Transactions on PAMI, 2007, 29(8):1454-1464.
- [103] Quan Y, Xu Y, Sun Y. A distinct and compact texture descriptor[J]. Image and Vision Computing, 2014, 32(4):250-259.
- [104] Beck J. Effect of orientation and of shape similarity on perceptual grouping[J]. Perceptual Psychophysics,1966,1(7):300-302.
- [105] Bergen J R, Adelson E H. Early vision and texture perception using feature distribution[J]. Pattern Recognition, 1999,32(3):447-486.
- [106] Tuceryan M, Jain A K. Texture segmentation using Voronoi Polygons[J]. IEEE Transactions on PAMI, 1990,12:211-216.
- [107] Carlucci L. A formal system for texture languages[J]. Pattern Recognition, 1972,5(1): 53-72.
- [108] Lu S Y, Fu K S. A Syntactic Approach to Texture[J]. Analysis. CGIP, 1978,7:303-330.
- [109] Yokoyama R, Robert M Haralick. Texture Pattern Image Generation by Regular Markov Chain[J]. Pattern Recognition, 1979,11:225-234.
- [110] Crivelli T, Cernuschi-Frias B, Bouthemy P, et al. Motion Textures: Modeling, Classification, and Segmentation Using Mixed-State Markov Random Fields[J]. SIAM Journal on Imaging Sciences, 2013, 6(4):2484-2520.
- [111] Liu L, Fieguth P W. Texture classification from random features[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3):574-586.
- [112] Geman S, Geman D. Stochastic relaxation Gibbs distribution and the Bayesian restoration of images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1984,16:721-741.
- [113] Cohen F, Fan Z, Attali S. Automated inspection of textile fabrics using textural models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1991, 13(8):803-809.
- [114] Mao J, Jain A K. Texture classification and segmentation using multiresolution simultaneous autoregressive models[J]. Pattern Recognition,1992,25(2):173-188.
- [115] Luo J, Savakis A E. Self-supervised texture segmentation using complementary types of features[J]. Pattern Recognition, 2001, 34(11): 2071-2082.
- [116] Bennett J, Khotanzad A. Modeling textured images using Generalized Long Correlation Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(12):1365-1375.
- [117] Rani M, Aggarwal S. Fractal Texture: A Survey[J]. Advances in Computational

- Research, 2013.
- [118] Lopes R, Dubois P, Bhouri I, et al. Local fractal and multifractal features for volumic texture characterization[J]. Pattern Recognition, 2011, 44(8):1690-1697.
  - [119] Liu F, Picard R W. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(7):722-733.
  - [120] Zhang J, Tan T. Brief review of invariant texture analysis methods[J]. Pattern Recognition, 2002, 35(3):735-747.
  - [121] Assefa D, Mansinha L, Tiampo K F, et al. Local quaternion Fourier transform and color image texture analysis[J]. Signal Processing, 2010, 90(6):1825-1835.
  - [122] Livens S, Scheunders P, Van de Wouwer G, et al. Wavelets for texture analysis[C] // an overview, Sixth International Conference on IET, 1997, 2:581-585.
  - [123] 安志勇, 曾智勇, 赵珊, 等. 基于纹理特征的图像检索[J]. 光电子·激光, 2008, 19(2):230-232.
  - [124] 钟桦, 杨晓鸣, 焦李成. 基于多分辨共生矩阵的纹理图像分类[J]. 计算机研究与发展, 2012, 48(11):1991-1999.
  - [125] Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8):837-842.
  - [126] Riaz F, Silva F B, Ribeiro M D, et al. Invariant gabor texture descriptors for classification of gastroenterology images[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(10):2893-2904.
  - [127] Lee Y H, Kim B, Rhee S B. Content-based image retrieval using spatial-color and Gabor texture on a mobile device[J]. Computer Science and Information Systems, 2013, 10(2):807-823.
  - [128] Riaz F, Hassan A, Rehman S, et al. Texture Classification Using Rotation-and Scale-Invariant Gabor Texture Features[J]. IEEE Signal Processing Letters, 2013, 20:607-610.
  - [129] Ojala T, Pietikäinen M, Hardwood D. A comparative study of texture measures with classification based on feature distribution[J]. Pattern Recognition. 1996, 29:51-59.
  - [130] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns[J]. IEEE Transactions on PAMI, 2002, 24(7):971-987.
  - [131] Liao S, Chung A C S. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude[J]. ACCV, 2007:672-679.
  - [132] Nanni L, Lumini A, Brahnam S. Local binary patterns variants as texture descriptors

- for medical image analysis[J]. Artificial intelligence in medicine, 2010, 49(2):117-125.
- [133] Liao S, Zhu X, Lei Z, et al. Learning multi-scale block local binary patterns for face recognition[C] // International Conference on Biometrics (ICB), 2007:828-837.
- [134] Wolf L, Hassner T, Taigman Y. Descriptor based methods in the wild[C] // Faces in Real-Life Images Workshop in European Conference on Computer Vision (ECCV), 2008:1-14.
- [135] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions[J]. IEEE Transactions on Image Processing, 2010, 19(6): 1635-1650.
- [136] Ren J, Jiang X, Yuan J. Noise-resistant local binary pattern with an embedded error-correction mechanism[J]. IEEE Transactions on Image Processing, 2013, 22(10): 4049-4060.
- [137] Keramidas E, Iakovidis D, Maroulis D. Fuzzy binary patterns for uncertainty-aware texture representation[J]. Electronic Letters on Computer Vision and Image Analysis, 2011, 10(1):63-78.
- [138] Kylberg G, Sintorn I M. Evaluation of noise robustness for local binary pattern descriptors in texture classification[J]. EURASIP Journal on Image and Video Processing, 2013(1):17.
- [139] Heikkilä M, Pietikäinen M, Schmid C. Description of interest regions with local binary patterns[J]. Pattern Recognition, 2009, 42(3):425-436.
- [140] 毋小省. 基于纹理谱特征的图像检索技术研究[D]. 河南: 河南理工大学, 2010.
- [141] Liao S, Law M W K, Chung A C S. Dominant local binary patterns for texture classification[J]. IEEE Transactions on Image Processing, 2009, 18(5):1107-1118.
- [142] Chao Zhu, Charles-Edmond Bichot, Liming Chen. Image region description using orthogonal combination of local binary patterns enhanced with color information[J]. Pattern Recognition, 2013, 46(7):1949-1963.
- [143] Zhao Y, Huang D S, Jia W. Completed local binary count for rotation invariant texture classification[J]. IEEE Transactions on Image Processing, 2012, 21(10):4492-4497.
- [144] 毋小省, 孙君顶. 基于局部边缘二值模式的图像检索[J]. 光电子·激光, 2013, 24(1):184-189.
- [145] Abdesselam A. Improving local binary patterns techniques by using edge information[J]. Lecture Notes on Software Engineering, 2013, 1(4):360-363.
- [146] Zhong F, Zhang J. Face recognition with enhanced local directional patterns[J]. Neurocomputing, 2013, 119(7):375-384.
- [147] 毋小省, 孙君顶. 基于改进方向纹理谱特征的图像检索[J]. 光电子·激光, 2012,

- 23(4):812-818.
- [148] Guo Z, Zhang L, Zhang D. A completed modeling of local binary pattern operator for texture classification[J]. IEEE Transactions on Image Processing, 2010, 19(6): 1657-1663.
  - [149] Sapkota A, Boulton T E. GRAB: Generalized Region Assigned to Binary[J]. EURASIP Journal on Image and Video Processing, 2013, 2013(1):35.
  - [150] Zhenhua Guo, Lei Zhang, David Zhang. Rotation invariant texture classification using LBP variance (LBPV) with global matching[J]. Pattern Recognition, 2010, 43(3):706-719.
  - [151] 毋小省. 改进的旋转不变区域纹理谱描述符[J]. 光电子·激光, 2011, 22(5):783-787.
  - [152] Pietikäinen M. Computer vision using local binary patterns[M]. Springer, 2011.
  - [153] Brahmam S, Jain L C, Lumini A, et al. Local Binary Patterns: New Variants and Applications[M]. Springer Berlin Heidelberg, 2014.
  - [154] <http://www.cse.oulu.fi/CMV/Research/LBP>.
  - [155] Junding Sun, Guoliang Fan, Xiaosheng Wu. New local edge binary patterns for image retrieval[J]. ICIP, 2013.
  - [156] Junding Sun, Guoliang Fan, Liangjiang Yu, Xiaosheng Wu. Concave-Convex Local Binary Features for Automatic Target Recognition in Infrared Imagery[J]. EURASIP Journal on Image and Video Processing, 2014.
  - [157] Chao Zhu, Charles-Edmond Bichot, Liming Chen. Image region description using orthogonal combination of local binary patterns enhanced with color information[J]. Pattern Recognition, 2013, 46(7):1949-1963.

## 基于压缩域的图像检索技术

随着压缩图像的使用越来越普遍和广泛，基于压缩域的检索技术得到了广泛关注和研究。其基本原理是通过挖掘图像压缩时的中间结果或最终码流中包含的信息，力争在不解码或部分解码的情况下提取表征图像内容的特征，并以此作为索引实现基于内容的图像检索。本章从常用图像压缩算法入手，系统介绍了近年来国内外学者的一些研究成果，最后详细论述了DCT压缩域中两种特征提取算法。

### 3.1 概 述

前面讨论的图像特征，主要考虑的是使用直接采集的原始格式的图像，而对于普遍存在的压缩形式的图像数据，必须先解压缩然后才能进行检索用特征的提取。对原始数据的分析处理，不但计算量大，而且需占用较多的中介存储空间，极不利于在计算资源有限的环境里进行，这无疑将影响图像检索系统的实用性和灵活性。特别是对一些大型的图像检索系统和要求实时性的动态检索系统，这种传统的处理模式往往难以满足要求。在这样的背景下，压缩域图像检索技术的研究受到了广泛的重视。在传统的图像检索技术中，操作对象是原始数据（可以直接获取的或者是对压缩数据进行解码以后得到的），而在基于压缩域的图像检索中，则试图直接在压缩数据上进行操作，不需要或不完全需要解压缩的环节。事实上，由于传统图像检索技术与压缩技术相互独立，现有的压缩编码算法所形成的压缩码流只考虑了存储和传输的需要，即在可视质量的允许下尽一切可能用最少的比特数来表征图像信息，而未考虑后续分析处理的需要，因此并不具有支持图像检索的能力。

基于压缩域的图像检索技术，实际上是把图像的压缩技术与检索技术融合在一

起,克服了现有的压缩技术与检索技术相分离所带来的局限性,力图实现快速、高效、灵活的检索技术。基于压缩域的图像检索技术与传统的基于原始域或解压域的图像检索技术相比有许多优点<sup>[1]</sup>。

(1) 在压缩域上检索可省略解压缩的附加环节,既可减少处理时间,也可减少设备开销。

(2) 许多图像压缩算法已对图像进行了大量的处理和分析,在检索中利用这些处理和分析的结果,可减少计算量,提高检索效率。

(3) 压缩域上的数据量比原始域或解压域上的数据量要少,这有利于提高整个系统的效率,尤其是在检索系统要求实时响应的场合。

(4) 基于特征的图像检索方法在存储图像的时候(建库)除了存储图像外还要存储相应的特征向量,而在压缩域上,某些特征向量的信息已经包含在压缩系数中,所以额外的存储量可以省去。

因此,如果能将压缩域提取的某些特征用 MPEG-7 所规定的特征描述符表示,则其检索技术更具有广泛意义;另外,由于对计算资源的需求减少,基于压缩域的图像检索技术非常有利于手持网络终端等计算资源有限的环境,对网络信息安全也具有重要作用。目前,基于压缩域的图像检索技术的研究已经引起了国内外的广泛关注,成为一个国际性研究热点<sup>[2]</sup>。

### 3.1.1 图像压缩技术

#### 1. 图像压缩机制的种类

图像数据的高效压缩和有效检索是构建多媒体信息业务和应用所必须解决的两个关键问题。对压缩域的图像进行检索必须根据其所采用的压缩技术而选择合适的检索技术。对图像压缩来说,任何压缩机制的基本思想都是去除数据中存在的冗余性。图像的压缩机制可以分为两种:有损压缩和无损压缩。

(1) 无损压缩也称信息保持编码,利用数据的统计特征进行压缩,解码图像和压缩编码图像严格相同,没有任何失真和信息损失。从数学上讲是一种可逆运算,但压缩率受到数据统计冗余度的理论限制,压缩率一般为 2:1~5:1。无损压缩广泛应用于文本数据、程序和特殊应用场合的图像数据的压缩。

(2) 有损压缩也称失真度编码,利用人类视觉对图像中的某些频率成分不敏感的特性,允许压缩过程中损失一定的信息,从而以一定的失真换来高的压缩率。有损压缩广泛应用于语音、图像和视频数据的压缩。

#### 2. 无损压缩

传统的压缩编码以 Shannon 信息论为出发点,用概率统计模型来描述信源。

Shannon 编码定理指出,在不产生任何失真的前提下,通过合理的编码,对每一个信源符号分配不等长的码字,平均码长可以任意接近于信源的熵。在这个理论框架下,出现了几种无失真信源编码方法,如行程编码(Run Length Coding, RLC)、霍夫曼编码(Huffman coding)、字串表压缩方法[也称 LZW (Lemple-Ziv&Welch) 压缩方法]、算术编码(Arithmetic Coding, ARC)等。这些信源编码方法通常称为熵编码(entropy coding),这种无失真编码的压缩率是很有限的,对于较复杂的自然图像,压缩率一般不超过 2:1。显然,无失真熵编码由于受压缩率的限制,使其难以满足大多数应用场合的要求<sup>[3]</sup>。

### 1) 行程编码

行程编码是一种相对简单的编码方案,是指在每一行扫描的像素中比较相邻像素的幅度(如高度等),当幅度有一个显著变化时,就说有一个行程存在,像素的连续长度和终点位置标记是其重要的参数。根据终点位置标记方法的不同,行程编码可以分为两类:行程终点编码和行程长度编码。前者的终点编码位置由扫描行的起始点算起至行程终点位置的像素数确定,而对后者来说,某行程长度的终点位置由它距前一终点的距离位置来确定。行程终点编码又分为两种:①线性码法,根据不同行程长度赋予不同的码字,大行程码字长,小行程码字短;②对数码法,它的码字长与行程长度的对数成正比。行程编码的码字结构相对较简单,适用于二值图像(如传真)的编码,因为 1 和 0 总是交替出现,对于不同串长度按其发生概率不同分配以不同码字,可以将 1 值的长度和 0 值的长度单独编码,也可将二者长度混合编码。行程编码的优点是非常适于大面积、相同码值较多的情况,缺点是对误码很敏感,一个值传输或存储出错,可能导致整幅图像混乱,而且对不连续的情况编码效果很差,甚至可能出现编码后的代码数比编码前还多。

### 2) 霍夫曼编码

霍夫曼编码是一种长度不均匀、平均码率可以接近信息源熵值的一种编码。它的基本编码思想是对于出现概率大的信号采用短字长的码,对于出现概率小的信号采用长字长的码,以达到缩短平均码长,从而实现数据压缩的目的。霍夫曼编码的最高压缩率可达到 8:1,但是在一般压缩过程中,很难达到这种压缩率。若图像中存在某个拥有长行程的字节值时,使用行程编码压缩方式可能会更好<sup>[4]</sup>。近几年发展起来的 Rice 编码是一种特殊的霍夫曼编码方法,它在编码时不需要码表,但却能提供与使用多个霍夫曼码表相同的功能, Rice 编码的速度优于广泛采用的算术编码,且具有接近于算术编码的压缩率,因而该编码方法在无失真图像压缩领域引起了重视<sup>[5]</sup>。目前,行程编码、霍夫曼编码和算术编码已被 JPEG、MPEG 等压缩标准所普遍采纳,用于对变换编码、预测编码之后的图像系数在无失真的前提下进一步处理以提高压缩率。



### 3) LZW 压缩方法

目前广泛采用的 LZW 压缩方法有两种类型。一种方法是在数据压缩过程中,寻找当前等待进行压缩处理的数据串是否在已经处理的数据串中出现过,如果曾经出现过,则利用指向该数据串的指针代替当前进行压缩的数据串,此时,字典是隐式的,它用曾经处理过的数据描述。另一种方法是为输入数据创建一个短语字典,如果在当前等待进行压缩的数据流中发现在字典中已经存在相应的短语,则利用该短语的相应索引取代原始数据。LZW 压缩方法的特点是压缩率很高,但比较复杂,不仅可以用于文字数据的压缩,还可以成功地用于某些图像的压缩处理。

### 4) 算术编码

算术编码压缩方法与霍夫曼编码压缩方法相似,都是利用比较短的代码取代图像数据中出现比较频繁的数据,而利用比较长的代码取代使用频率比较低的数据,从而达到数据压缩的目的。它同时又采用了字典压缩编码的思想,不仅压缩数据值,而且压缩值序列,从而达到更理想的压缩率,尤其适合大多数数据由相同的重复序列组成的图像文件。其基本思想是将每个不同的序列按照出现的频率映射到 0 和 1 之间相应的数字区域内,该区域表示成可以改变精度的二进制小数,其中出现频率越高的数据利用精度越高的小数表示。算术编码算法可以大幅度地减小文件长度,压缩率甚至可以达到 100:1,在 JBIG 与 JPEG 图像文件格式的数据处理中占据很重要的地位。由于压缩算法比较复杂,其同时还受到几项 IBM 的专利保护,从而导致了算法许可性的不确定性<sup>[6]</sup>,阻碍了算术编码压缩算法的推广。针对不同的图像文件,算术编码的压缩率主要与源文件的数据分布及其标准模式的精度有关。

## 3. 有损压缩

除了无失真熵编码外,还有一些实现信源冗余压缩的方法,如预测编码(predictive coding)、变换编码、矢量量化等传统的编码技术。这类编码技术采用的方法是去除图像数据中的冗余信息或对图像内容表征不太重要的细节分量以尽可能少的码字来表示要处理的图像,属于有损压缩方法。

### 1) 预测编码

预测编码是研究最早的图像压缩方法,它的基本思想是利用图像数据所具有的空间或时间相关性,用相邻的已知像素(或图像块)来预测当前像素(或图像块)的取值,然后对预测误差进行量化和编码。预测编码的关键在于预测算法的选取,这与图像信号的概率分布有很大的关系,实际应用中常根据大量的统计结果采用简化的概率分布形式来设计最佳的预测器,有时还使用自适应预测器来刻画图像信号的局部特性,以提高预测效率。预测编码方法计算简单,但由于是基于差值信号的统计特性发展起来的,故存在着一些缺点。例如,对黑白灰度有突变的点,会有较大的预测误差,

使重建图像的边缘模糊,分辨率降低;对图像亮度变化缓慢的区域,其差值信号为零,但因预测值偏大而使重建图像产生噪声。

## 2) 变换编码

变换编码是消除图像数据空间相关性的一种更为有效的图像压缩方法,它通过正交变换实现数据压缩。图像经过正交变换后,经过多维坐标系中适当的坐标旋转和变换,能够把分布在各个坐标轴上的原始图像数据,在新的适当的坐标系中集中到少数坐标轴上,因而只需对少数的高能量系数进行适当的量化和熵编码,就可以用较少的编码比特数来表示一幅图像,实现图像的压缩编码。变换编码实质上是通过对变换使图像数据在变换域上最大限度地不相关,图像经过变换后,系数的空间分布和频率特性可能会与人眼的视觉系统特性相匹配,可以进一步利用人眼视觉系统的生理和心理特点来设计需要的编码系统<sup>[7]</sup>。从数学的角度看,可用于图像变换编码的正交变换除傅里叶变换、Walsh-Hadamard 变换外,还有离散余弦变换(DCT)、哈尔变换、K-L 变换等,不同的变换有不同的压缩效果(压缩率和重建图像质量)。目前,最流行的静态图像压缩标准 JPEG 采用的是 DCT,一方面是因为 DCT 具有去相关性和能量压缩特性,另一方面是因为 DCT 具有快速实现算法。然而,基于 DCT 的变换编码的主要不足在于变换方法不具有好的时频局域性和全局变换,因而在提高编码压缩率时会出现块效应,导致图像的边缘轮廓模糊,严重影响重建图像的主观质量。而且,现代传输媒体、互联网等要求图像编码传输能够提供关于质量、分辨率等可扩展结构的要求,但这些灵活性要求与 DCT 编码的结构很难实现有机的结合。

## 3) 矢量量化

矢量量化(Vector Quantization, VQ)是另一种高效的图像压缩技术。根据 Shannon 信息率失真理论,只要码书数足够大,矢量维数足够大,矢量量化方法能够得到 Shannon 定理所规定的下限<sup>[8]</sup>。实际的矢量量化中编码端和解码端具有相同的码书,码书由所有可能矢量值集合的有序子集组成,具体编码时,编码端依据选定的距离测度(或称代价函数)在码书中对输入的图像分块矢量进行匹配,然后对匹配码的码字序号进行编码,从而实现由一个矢量所需比特数到一个码字序号所需比特数的压缩。矢量量化方法实现的关键在于最佳码书的设计和快速匹配算法的选择,码书性能的好坏直接影响着整体矢量量化的性能。近年来,许多学者利用一些新兴理论,如神经网络和模糊理论等,对码书的设计算法进行了深入的研究,提出一些比传统码书设计算法性能更优的算法,如竞争性学习算法<sup>[9]</sup>、模糊竞争性学习算法<sup>[10]</sup>等。其中,竞争性学习算法可以说是传统 LBG 算法的自适应形式,因而它仍保留 LBG 算法的一些缺陷,如存在码矢欠利用问题。为此,人们进一步将模糊理论引入竞争性学习算法中,提出了模糊竞争性学习算法。模糊竞争性学习算法中每个权矢量都按照隶属度函数得到不同程度的调整,较充分地利用了输入训练矢量与码矢之间的距离信息来修正模糊隶属

度函数,使之有效地控制码矢的训练调整过程,从而较好地解决码矢欠利用等问题。

#### 4) 金字塔编码方案

金字塔编码方案是把图像分解成许多不同分辨率的子图像,并把高分辨率(尺寸较大的)的子图像放在下层,把低分辨率(尺寸较小的)的子图像放在上层,从而像一个金字塔,借助拉普拉斯金字塔算法,对图像的每一层分别量化、编码,并对视觉不敏感的层进行粗化,用较少的码字编码,从而达到压缩的目的。解码是逐层累积,从轮廓到细节重建图像。这样的等级图像结构特别适用于检索型的应用场合,并且可以根据需要给不同的塔层赋予不同的优先级。图像的金字塔表示方法可以用很紧凑的方法来进行图像的编码,但它的缺点在于编码过程中会增加一定的数据量,因而无法达到高的压缩率。

### 4. 编码技术的新发展

以上的几种压缩方法在20世纪80年代中期以前的研究中占据着重要的地位,其编码实体主要是像素或像素块,称为第一代编码技术。目前已得到普遍应用的JPEG、MPEG和H.261等压缩编码国际标准均采用这种技术。但是,第一代编码技术并未考虑信息接收者的主观特性,也不关心图像信息的具体含义和重要程度,只是力图去除数据冗余,因此是一种低层次的编码技术。到了20世纪80年代中后期,相关学科的迅速发展和新兴学科的不断涌现为图像压缩编码的研究和发展注入了新的活力。许多学者结合模式识别、计算机图形学、计算机视觉、神经网络、小波变换和分形几何等理论对图像压缩编码的新技术进行了探索,同时关于人类视觉生理、心理特性的研究成果也为图像压缩提供了新的思路 and 许多新的方法,称为第二代编码技术,如基于分形的图像压缩方法、基于模型的图像压缩方法及基于小波变换的图像压缩算法<sup>[10]</sup>等。第二代编码技术充分考虑了人眼视觉系统特性的影响,认为人眼是图像的最终接收者,关心的是如何去除图像内容的冗余。

#### 1) 子带编码

子带编码是第一代编码技术和第二代编码技术的过渡。它的基本思想是把图像信号通过一组带通滤波器分解成不同频带内的分量,然后在每个独立的子带中,对信号进行降速采样和单独编码。首先将图像进行滤波而产生一个图像集合,其中每个都占据一个空间频率的限制范围,这些图像分量就称为子带。由于子带与原来满频带图像相比,均占有一个小的带宽,就可对它们进行降速采样,这种滤波处理和亚抽样称为图像的分解阶段。然后利用一个或多个编码器对子带进行编码,各子带图像的统计特性是以人眼对它的敏感程度来确定的,可以根据信息论和人眼的视觉特性,利用不同的比特率甚至是不同的编码技术对各子带编码。对子带编码序列的译码,则要使用升速取样及适当滤波而获得重建的子带图像,然后将各子带相加,

最后得到完整的重建图像。子带编码的工作原理如图 3.1 所示。在整个编码过程中，构成子带并不带来任何压缩效果，这是因为全部子带所需的样本数与图像所需的样本数相同。子带编码的积极意义在于，这些子带可以较原图像有更高的编码效率，即子带编码的高效不在于形成子带，而在于各子带的高效编码。子带编码的优点是由于量化在各子带内单独进行，量化噪声便被限制在各子带内，可以防止能量较小的频带内信号受其他频带内量化噪声的干扰和影响。

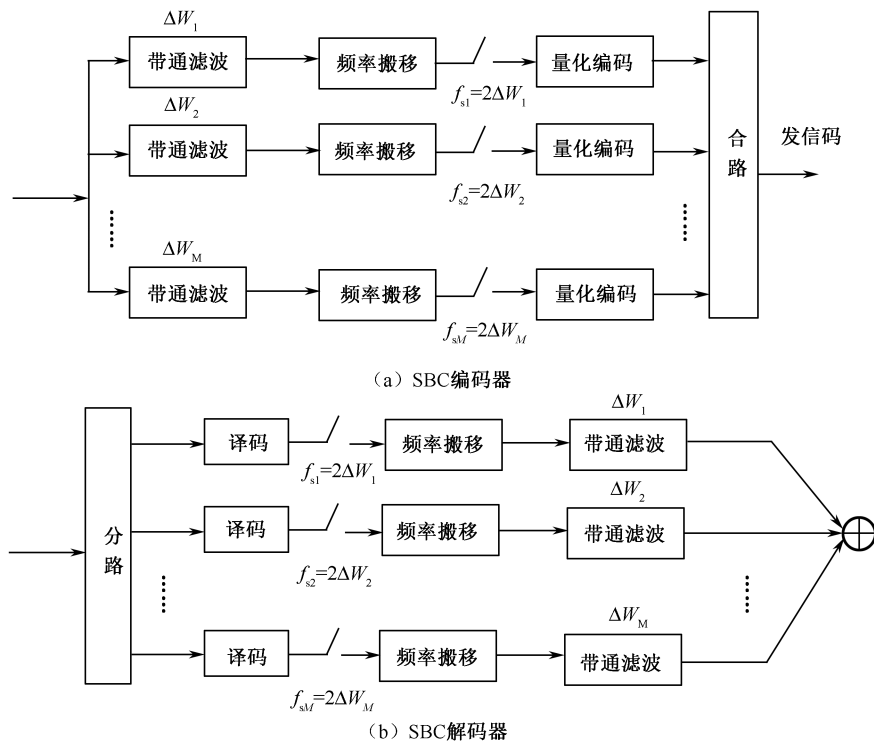


图 3.1 子带编码的工作原理

## 2) 分形编码

分形编码 (fractal coding) 是在分形几何理论上发展起来的一种编码方法。Barnsley M. 最早引入了迭代函数系统来刻画图像中的自相似性，并将其用于图像压缩，对某些特定图像获得了极高的压缩率，但是他的方法需要人工干预。为此，Jacquin 提出了基于分块迭代函数系统的自动分形编码，从而将分形在图像编码上的应用推进了一大步<sup>[11]</sup>。分形编码的关键在于图像 IFS (Iterative Function System) 码的获取，目前对某些图像可获得 30~70 倍的压缩率。然而，分形编码的理论基础决定了它只有对那些明显具有自相似性或统计自相似性的图像才有较高的压缩率，对一般的图像，特别是相似性不强的图像，基于分形的方法其效率并不高；另外，分形编码是一种非对

称方法（编码复杂度远远高于解码复杂度），其实现的结构很难利用人眼的视觉特性，只有和其他编码方法相结合，才能得到较高的编码效率和较低的实现复杂度<sup>[12]</sup>。

### 3) 模型编码

模型编码（model-based coding）是一种被认为很有前景的低比特率编码方法。它利用计算机视觉和计算机图形学中的方法和理论，在编码、解码两端分别建立了相同的模型，在编码端对输入图像进行分析，获取模型参数，并将模型参数传递给解码端；解码端根据接收到的模型参数进行图像合成，重建原始图像。因此，模型编码的核心在于模型的建立和模型参数的获取<sup>[13]</sup>。

### 4) 小波图像压缩方法

小波图像压缩方法是近年来受到广泛重视并已逐步得到应用的变换编码技术。与传统的 DCT 相比，离散小波变换（DWT）具有良好的时频局部化特性和较好的能量集中特性；同时由于小波变换是对整幅图像进行的，因此基于小波变换的图像压缩方法能够很好地克服方块效应；此外，小波变换的多分辨率特性提供了利用人眼各种视觉特性的良好机制。正是由于这些原因，众多学者对基于小波变换的图像压缩方法进行了研究。Shapiro J.M. 利用小波分解不同频带之间的关联，采用访问演出树实现了自嵌套的图像压缩方法，在重建图像质量良好的条件下，获得了 128 倍的高压缩率<sup>[14]</sup>。而零树编码所获得的高性能使得研究者们对其进行了许多改进，当前，基于小波零树思想的小波图像压缩方法仍被认为是利用小波变换对图像进行压缩编码所获得的最好结果。因此，新一代的多媒体压缩标准 MPEG-4 中静态纹理的编码就是采用了基于零树思想的小波编码方法，而新一代静止图像的压缩标准 JPEG2000 也全面采用基于小波变换的压缩编码算法。然而，就目前小波压缩方法所获得的压缩率来看，基于小波变换的图像压缩方法所具有潜能还远远未发挥出来，在最佳小波变换基函数的选择、人眼视觉特性的利用、变换系数的有效组织及与其他编码方法的有效结合等方面仍需进一步深入的研究。

## 3.1.2 静态图像压缩标准

自 Oliver 提出电视信号的线性 PCM 编码理论以来，图像压缩编码经历了半个多世纪的发展，已进入了广泛应用和深入研究的高速发展时期，其标志就是几个关于图像编码的国际标准的制定。

### 1. G3 和 G4

这两个标准最初是为传真应用而设计的，现也用于其他方面。G3 采用了非自适应、

1D 游程编码技术。对每组  $N$  行 ( $N=2$  或  $N=4$ ) 扫描线中的后  $N-1$  行也可以用 2D 方式编码。G4 是 G3 的一种简化版本, 其中只使用 2D 方式编码。实验表明, G3 对它们的压缩率约为 15:1, G4 的压缩率一般比 G3 要高一倍<sup>[15]</sup>。

## 2. JBIG

由于 G3 和 G4 是基于非自适应技术的, 对半调灰度图像编码时常产生扩展 (而不是压缩) 的效果。ISO/IEC 和 ITU (原 CCITT) 联合委员会于 1992 年提出和开发 JBIG (Joint Bi-level Image Experts Group) 标准, 这是适用于二值和低精度灰度 (低于 6bit/像素) 图像的无损压缩标准<sup>[16]</sup>, 为区别于后续制定的标准, 简称 JBIG1。

JBIG1 采用 2D 常规或自适应预测模式、自适应算术编码等方法, 能自适应图像的特征, 对不同的图像类型均具有稳健性, 因而具有较高的编码效率。它能满足标准化的要求, 对于那些满足二值图像质量的用户, JBIG1 能通过减小文件大小来节约巨大的费用开支, 能更快地传输图像文件, 能在相同的文件大小条件下提供更好的图像质量, 能改善图像质量与大小之间的矛盾。由于在获取灰度图像时, JBIG1 结构并不仅限于标准的 8bit, 因此能提供较小的灰度级文件; 又因为 JBIG1 将灰度数据存储于独立的位平面中, 因此能高效地被应用在任何比特数的灰度数据上。JBIG1 编码器的总体框架如图 3.2 所示<sup>[17]</sup>。

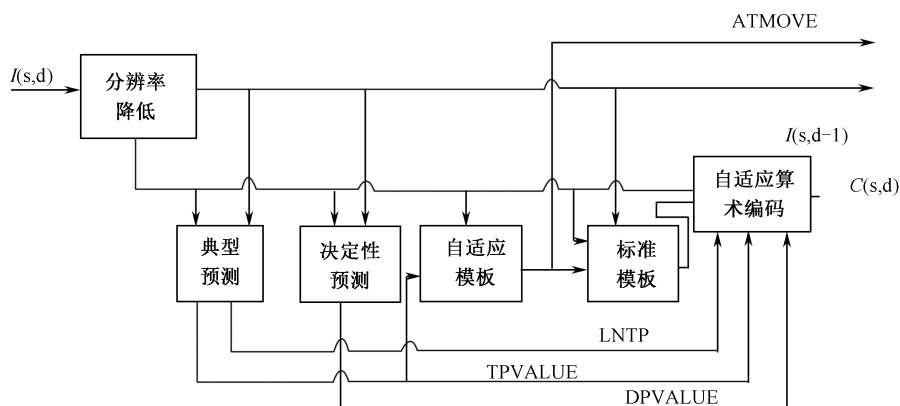


图 3.2 JBIG1 编码器的总体框架

由于采用了自适应技术, JBIG1 的编码效率比 G3 和 G4 要高。对于打印字符的扫描图像, 压缩率可提高 1.1~1.5 倍; 对于计算机生成的打印字符图像, 压缩率可提高约 5 倍; 对于用抖动或半调表示的灰度图像, 压缩率可提高 2~30 倍。虽然 JBIG1 也有有损压缩的能力, 但其有损压缩的图像质量明显低下。1999 年 7 月, 该小组又制定了 JBIG2 标准, 支持有损、无损和渐进编码, 其设计目标是让无损压缩的性能超过已有的其他标准, 让有损压缩在取得比无损压缩更高的压缩率的情况下, 具有几乎不可见的质量下降<sup>[18]</sup>。

### 3. JPEG

JPEG (Joint Photographic Experts Group) 是联合图像专家小组的简称。1991 年, 该小组提出 ISO CD10918 标准建议草案, 1992 年, 该标准成为国际标准 ISO/IEC IS10918, 后来将该标准称为 JPEG, 它主要涉及连续色调 (灰度和彩色) 静止图像的压缩编码。JPEG 标准支持两种图像模式, 即顺序型和渐进型, 用以满足用户对具体应用的不同需要。JPEG 压缩算法分为两大类: 无失真压缩和有失真压缩。使用无失真压缩算法将原图像数据转变为压缩数据, 该压缩数据经过对应的解压缩算法处理后可以获得与原图像完全一致的重建图像; 有失真压缩算法主要基于 DCT, 所生成的压缩图像数据经过解压缩生成的重建图像与原图像在视觉上保持基本一致, 压缩率越大, 视觉一致性越差。JPEG 共有 4 种编码模式: ①无失真模式, 基于空间预测的无损压缩算法, 可保证无失真地重建原始图像; ②顺序模式, 按从上到下、从左到右的顺序对图像进行编码, 也称为基本系统; ③渐进模式, 按由粗到细的顺序对一幅图像进行编码; ④分层模式, 以各种不同的分辨率对图像进行编码。

根据编码种类的不同将 JPEG 系统分成基本系统和扩展系统。基本系统由 DCT 顺序模型及霍夫曼编码组成, 所有符合 JPEG 标准的设备都具备基本系统; 扩展系统提供不同的选项, 即除基本系统外的其他编码模式, 如渐进型编码、算术编码、无失真编码、分层编码等。

基于 DCT 的 JPEG 基本系统利用人的视觉特性, 通过 DCT 系数量化和无损压缩编码去掉视觉不敏感信息, 并最终实现图像数据压缩的目的。编解码过程如图 3.3 所示<sup>[2]</sup>, 包括图像数据输入、基于 DCT 的编码和压缩图像数据流处理的输出 3 个步骤, 解码的过程沿着箭头相反的方向进行。

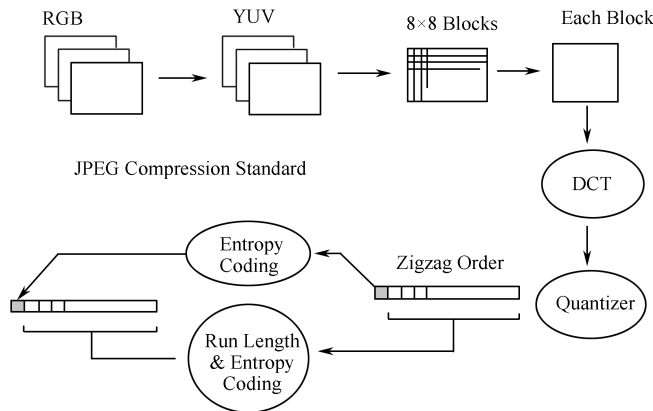


图 3.3 JPEG 压缩过程

### 1) 颜色空间转换

JPEG 算法与颜色空间无关, 编码原图像的输入, 可以是单色图像的灰度值, 也可以是彩色图像的亮度分量或色差分量信号。图像输入端首先将 RGB 空间转换到 YUV 空间, 亮度信号用  $Y$  保存, 色彩分量用两个颜色差值分量  $U$  和  $V$  来保存。

### 2) DCT

在对图像进行 DCT 时, 考虑到编码效率和运算复杂度, 采用了分块技术, 一般采用  $8 \times 8$  分块, 因此首先对图像进行  $8 \times 8$  分块, 再分别对每块进行 DCT。采样精度为  $p$  位, 把  $[0, 2^p - 1]$  范围内的无符号整数延拓为  $[1 - 2^{p-1}, 2^{p-1} - 1]$  范围内的有符号整数, 以此作为离散余弦正变换 (FDCT) 的输入, 即

$$F(u, v) = \frac{1}{4} C_u C_v \sum_{i=0}^7 \sum_{j=0}^7 f(i, j) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \quad (3-1)$$

在解码器的输出端, 离散余弦逆变换 (IDCT) 输出  $8 \times 8$  的数据块用以重构图像, 即

$$f(i, j) = \frac{1}{4} \left[ \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos \frac{(2i+1)u\pi}{16} \cos \frac{(2j+1)v\pi}{16} \right] \quad (3-2)$$

图像经 DCT 后, 输出 64 个 DCT 系数, 对应于  $u=0$ 、 $v=0$  的系数, 称为直流分量, 即 DC 系数, 其余 63 个系数称为 AC 系数, 即交流分量。DCT 实际上是像素域的低通滤波器, 其低频分量都集中在左上角, 且包含了图像的色彩、亮度等主要信息, 高频分量则分布于右下角, 代表图像的细节部分。图 3.4 给出了一幅  $8 \times 8$  的子图像块经过 DCT 后得到的变换系数。因为在一幅图像中像素之间的灰度和色差信号变化缓慢, 在  $8 \times 8$  子块中像素之间相关性很强, 所以经过 FDCT 后, 在低频集中了数值大的系数, 远离直流系数的高频交流系数则大多为零或趋于零, 这就为数据压缩提供了可能。

139	144	14	153	155	155	155	155	235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	13
144	151	15	156	159	156	156	156	-22.6	-18.5	-6.2	-3.2	-1.0	-0.1	0.4	-1.2
150	155	16	163	158	156	156	156	-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1
159	161	16	160	160	159	159	159	-7.1	-0.9	0.2	1.5	1.9	-0.1	0.0	0.3
159	160	16	162	162	155	155	155	-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3
161	161	16	161	160	157	157	157	1.8	-0.2	-1.6	-0.3	-0.8	1.5	1.0	-1.0
162	162	16	163	162	157	157	157	-1.3	0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8
162	162	16	161	163	158	158	158	-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

(a) 子图像块

(b) DCT 系数

图 3.4 子图像块及 DCT 系数



### 3) 量化

经 DCT 后, 其能量主要集中在变换域的左上角, 即频率系数较低的区域, 此时对变换域的系数值进行量化后, 许多高频区域已经为零, 特别有利于图像的压缩。这里的量化是根据人眼对频率的反映特性来设置的, 其目的是尽量去掉引起视觉冗余的数据并保证图像的视觉质量。在 JPEG 标准中采用线性均匀量化器, 对 64 个 DCT 系数  $F(u,v)$  除以量化步长, 四舍五入取整。其定义为

$$F^Q(u,v) = \text{Round} \left[ \frac{F(u,v)}{Q(u,v)} \right] \quad (3-3)$$

其中,  $Q(u,v)$  是量化步长。量化表元素随 DCT 系数的位置而改变, 与 64 个变换系数一一对应, 也是 DCT 编解码信息损失的根源。JPEG 压缩标准中采用的两个量化表实例如图 3.5 所示。

在接收端要进行逆量化, 逆量化的计算公式为

$$F^{Q'}(u,v) = F^Q(u,v) \cdot Q(u,v) \quad (3-4)$$

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

17	18	2	47	99	99	99	99
18	21	2	66	99	99	99	99
24	26	5	99	99	99	99	99
47	66	9	99	99	99	99	99
99	99	9	99	99	99	99	99
99	99	9	99	99	99	99	99
99	99	9	99	99	99	99	99
99	99	9	99	99	99	99	99

(a) 亮度量化表

(b) 色度量化表

图 3.5 JPEG 压缩标准中采用的两个量化表实例

### 4) 编码

对量化后的数据, 即 DC 系数和 AC 系数分别进行编码。DC 系数是  $8 \times 8$  子块 64 个采样的平均值。因为相邻的  $8 \times 8$  子块的相关性强, 相邻块的 DC 系数差值很小, 采用差分编码 (DPCM) 的方法, 对  $\Delta DC_i = DC_i - DC_{i-1}$  进行编码可以用较少的位数, 提高压缩率。经量化后的 63 个 AC 系数, 根据其系数的频率分布特点进行处理。由于低频部分反映了图像的主要内容, 通常具有较大的系数值, 而高频部分系数较小, 量化后零值较多。对所有的频率系数按 Zigzag 方式重新排列 (如图 3.6 所示), 使系数按照递增的空间频率定性地排列, 保证低频分量先出现, 高频分量后出现, 会出现成片的零值, 这时采用 RLE 编码的方式能有效地压缩成片的零值区域。

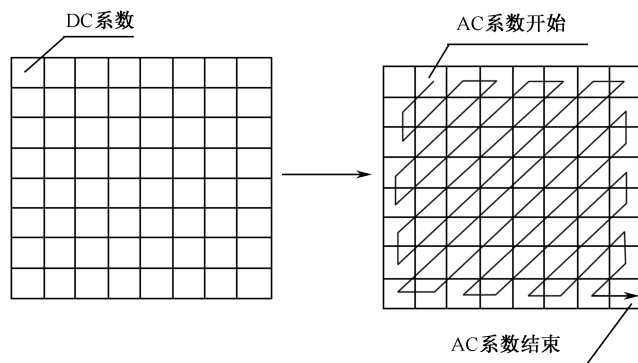


图 3.6 变换后的 DC 系数和 AC 系数及 AC 系数编码采用的处理方式

经过行程编码后，JPEG 标准对得到的数据再进行熵编码，旨在从统计的角度进一步降低数据量。使用熵编码可以对 DPCM 编码后的直流 DC 系数和行程编码后的交流 AC 系数进行进一步的压缩。JPEG 建议采用两种熵编码方法：霍夫曼编码和自适应二进制算法编码（adaptive binary arithmetic coding）。熵编码的过程大致可以看成由两个步骤组成，首先将重新排序并且经过行程长度编码之后的量化系数转换成为符号队列，然后将符号队列转换成数据流。符号的定义与之前的 DCT 编码及之后的熵编码操作无关，仅代表相应的数据，以供进行统计编码使用。熵编码从统计的角度进一步减少数据的冗余度，去掉了一些对视觉不重要的信息，减少了数据冗余，从而达到一定压缩的目的。

### 3. JPEG2000

随着图像种类和数据量的增加，JPEG 越来越不能满足实际应用的要求，其中包括压缩率、对多种压缩模式的选择和对不同种类图像的适应性等。JPEG2000 正是针对 JPEG 的缺陷而提出的一种静止图像压缩标准<sup>[19]</sup>。与传统的 JPEG 标准最大的不同在于 JPEG2000 标准扬弃了以 DCT 为主的块编码方法，而改采用以小波变换为主的多解析编码方法。作为 JPEG 标准的一个更新换代标准，它的目标是进一步改进目前压缩算法的性能，以适应低带宽、高噪声的环境，以及医疗图像、电子图书馆、传真、Internet 网上服务和安保等方面的应用。JPEG2000 规定了一系列对连续色调、二值、灰度或彩色数字静止图像的无失真或有失真编解码方法，JPEG2000 的基本编码框图如图 3.7 所示<sup>[20]</sup>。

在 JPEG2000 标准中，一幅图像首先被分解为若干方形的图像片，图像片是图像编解码的基本单位，各个图像片的编解码过程是独立进行的。分片的目的有两个：降低内存要求和实现特定区域的编解码。

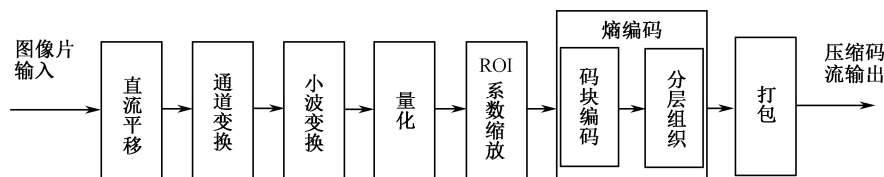


图 3.7 JPEG2000 的基本编码框图

### 1) 直流模块

该模块是可选模块，仅适用于图像分量的采样值为无符号数的情况，目的是去掉图像片中的直流分量，从而使小波变换后系数取正值和取负值的概率基本相等，提高后续的自适应熵编码效率。

### 2) 分量变换

该模块也是可选模块，仅针对多分量图像（如彩色图像）适用。JPEG2000 标准中定义了两种分量变换方式：可逆分量变换（Reversible Component Transformation, RCT）和不可逆分量变换（Irreversible Component Transformation, ICT）。RCT 和 ICT 都是去相关变换，通过对图像的 3 种分量进行相应的变换将图像变换到合适的色彩空间，提高整体编码效率。RCT 既适用于无损压缩，又适用于有损压缩。ICT 仅适用于有损压缩。

### 3) 小波变换

与 JPEG 相比，JPEG2000 最大的改进是以离散小波变换代替了离散余弦变换。DCT 作为准最优变换，在图像处理中占据很重要的地位，但它有个重要的缺点是不具有时频局域性，考察的是整个时域过程的频域特征或整个频域过程的时域特征。因此，对于平稳过程，DCT 有很好的效果，但对于非平稳过程，就存在明显不足。在图像压缩领域表现更为明显，对于细节丰富、频率变化大的图像，DCT 的压缩效果差；另外，也无法实现 ROI 编码。而小波变换的最大特点就是具有良好的时频局域性，既能考察局部时域过程的频域特征，也能考察局部频域过程的时域特征，并且可以在高频时考察窄的时域窗，而在低频时考察宽的时域窗。因此，不论是对于平稳过程还是对于非平稳过程，它都是强有力的工具，比 DCT 更适用于处理数字图像这样的非平稳信号。JPEG2000 标准中采用了两种小波变换滤波器组：有损变换（Daubechies 9/7 滤波器）和无损变换（Le Gall 5/3 滤波器）。这两种小波变换均可采用计算复杂度低的小波提升算法。

### 4) 量化

小波变换后虽然变换系数的个数并无减少（与原图像采样点个数相比），但信息

分布发生了很大的变化，大部分能量集中在少数的小波系数中，利用量化可大量减少幅度很小的系数所携带的能量，提高整体压缩率。量化主要是针对有损压缩进行的，量化的关键是根据变换后图像的特征及重构图像质量要求等因素选取合理的量化步长。关于有损压缩量化步长的选取，JPEG2000 标准中有很大的灵活，没有给出统一的方法，用户可根据实际应用自行选择设计，并不影响解码过程的进行，但需要把选取的量化步长作为参数放入比特流中传给解码器。量化步长与小波分解子带是一一对应的，一个子带一个量化步长，这样可充分利用人眼的视觉特性来提高编码效率。

### 5) 感兴趣区系数处理

该模块也是可选模块，用于对感兴趣区编码。JPEG2000 中对 ROI 编码所采取的策略是通过系数缩放使感兴趣区的数据位于更高的编码位平面（相对于背景区）。这样在编码比特流中，ROI 中的数据就排在背景区数据的前面。当压缩率较高时，可以优先保证 ROI 的恢复质量。

### 6) 熵编码

经分量变换、小波变换及量化后的图像数据，在一定程度上减少了空域或频域上的冗余度，但这些数据在统计意义上仍存在一定的相关性，为此采用熵编码来进一步消除数据间的统计相关。JPEG2000 熵编码过程可分为两个步骤：嵌入式码块编码和分层组织压缩位流。前者的基本思想是使压缩后的码流划分为若干逐级包含的子集，每一子集表示对原图像的单元压缩；后者则可以使压缩码流具有质量上的可分级性，实现图像的渐进式传输。

嵌入式码块编码的基本思想是使压缩后的码流划分为若干逐级包含的子集，每一子集表示对原图像的单元压缩，嵌入式码流可在任意处被截断，得到不同码率、不同质量的重构图像。首先将量化后的各子带小波系数划分成方形的码块，按图 3.8 所示把码块中的系统分解为若干个位平面<sup>[20]</sup>，以这些位平面为单位，从最高有效位平面开始对每个位平面的所有位进行算术编码。图中，C0、C1、C2、C3 等表示待编码码块

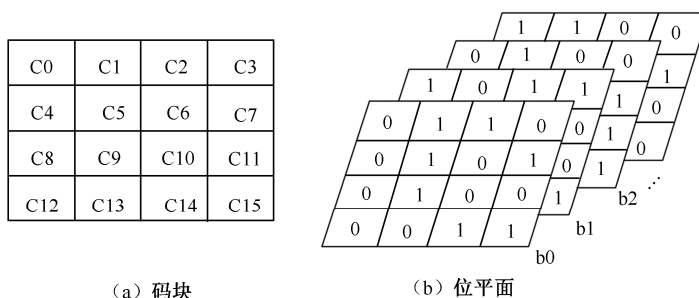


图 3.8 码块及其分解位平面示意图

中的各量化系数（假设码块大小为  $4 \times 4$ ），b0、b1、b2 等表示从高有效位到低有效位的位平面。

在此基础上，采用 PCRD（Post-Compression Rate-Distortion）优化算法，按照率失真最优原则，计算每一独立码块位流的截断点，根据给定的一系列压缩位率，可以把每一独立码块位流分割成若干不同长度的位流段，形成码块的嵌入式位流。嵌入式位流使压缩码流具有质量上的可分级性，从而可实现网络浏览、远程图像的渐进式传输。将所有码块位流按照截断点分层组织，形成所谓的质量层  $Q_q$ ，对于每个  $q$ ，所有编码块对  $Q_1$  到  $Q_q$  作的贡献表示原图像的一个率失真最优压缩，将压缩码流分层组织，每一层含有一定的质量信息，在有前面层的基础上，改善图像质量。码块的嵌入式位流分布在不同的层上，不同的码块对不同的层有不同的贡献，即使同一码块对不同层，或者不同码块对同一层，贡献也可能不同，有的码块甚至对某一层没有贡献。图 3.9 为码块对各层压缩位流的贡献示意图。

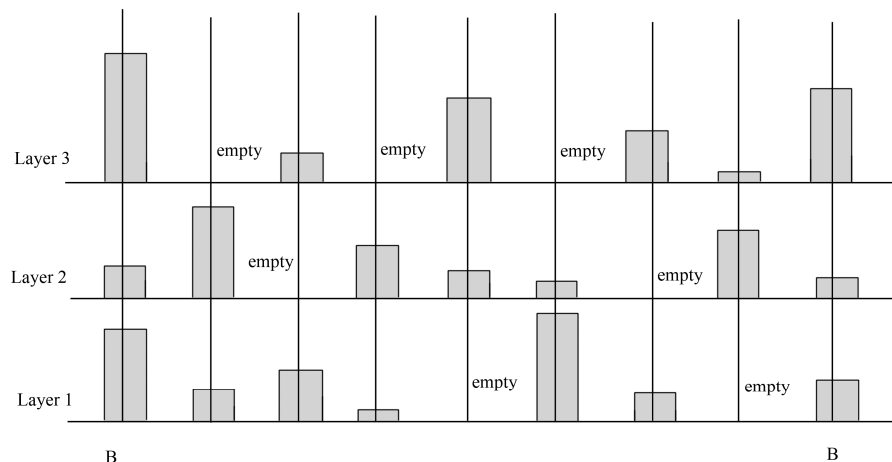


图 3.9 码块对各层压缩位流的贡献示意图

## 7) 打包

形成一定格式的压缩码流。为了适合图像交换，JPEG2000 标准规定了详细的语法结构来存放压缩码流及解码所需参数，以包为单位，形成最终码流。

当码率很低（高压缩率）时，或者对图像的质量要求非常高时，JPEG2000 的性能要优于 JPEG，对许多图像的测试表明，在压缩率大两到三倍的情况下，JPEG2000 编码造成的失真与 JPEG 造成的失真可以比拟；不过对无损或接近无损的压缩，JPEG2000 相对于 JPEG 的优势不大<sup>[21]</sup>。

### 3.1.3 压缩域图像检索的原理

基于压缩域的图像检索技术研究的关键问题是如何通过图像处理直接在压缩数据中提取图像的内容特征，即通过挖掘图像压缩时的中间结果或最终码流中包含的信息，力争在不解码或部分解码的情况下提取表征图像内容的特征，并以此作为索引实现基于内容的图像检索。这是一个全新的研究思路，它将图像的压缩技术与检索技术融合在一起，避免了全解压的额外操作，大大提高了检索系统的实时性、灵活性和有效性。基于压缩域的图像检索系统的基本组成如图 3.10 所示<sup>[22]</sup>，首先图像的特征可直接在压缩域中快速提取；其次，有利于动态图像数据的管理，不需要预先对图像数据进行处理（即提取特征，形成图像特征库）。因此，特别适用于 Internet、Web 数据库的检索。

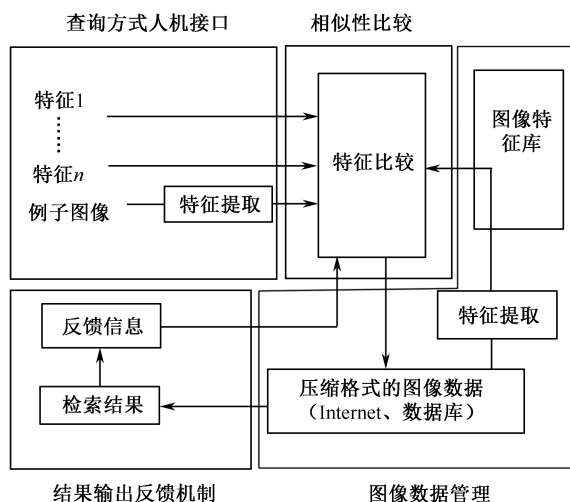


图 3.10 基于压缩域的图像检索系统的基本组成

面对图像压缩数据，传统的方法是将压缩图像进行全解码后在原始像素域进行处理，而在压缩域进行图像处理，就是在图像压缩码流不经解码或少量解码的情况下直接从图像压缩码流中提取图像的特征，即如何进行压缩域的图像处理，以得到图像的特征。图 3.11 给出了两种模式的比较示意图，可见基于压缩域的图像检索技术省去了一些中间环节。在压缩域的图像检索中，其提取数据的处理位置按压缩码流的可操作程度可分为 3 个位置。以变换编码的编解码过程为例，压缩域可操作的 3 个位置如下<sup>[22]</sup>。

(1) 熵解码前：这是压缩域处理的理想位置，但是经熵编码后所形成的压缩码流不具备结构化信息，其数据项并非字节对齐，不利于计算机的处理，所以，实际

上在此处能进行的针对图像的操作较少, 仅限于一些特殊情况, 如利用修改 JPEG 码流的量化表来达到图像锐化的目的。

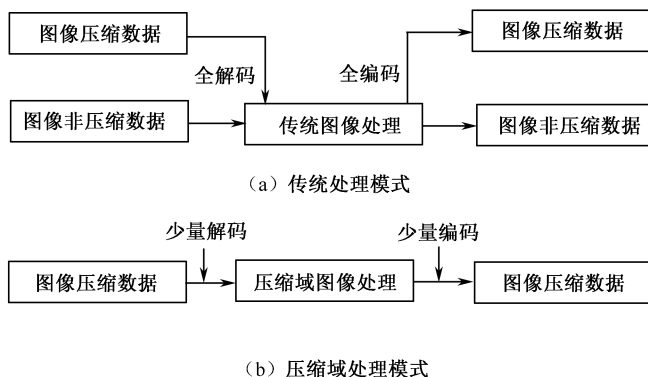


图 3.11 图像压缩数据的两种处理模式

(2) 熵解码后：由于直接在熵解码之前进行图像处理非常困难, 因此, 压缩域的处理操作一般要首先进行熵解码, 再进行处理。在图像的编码过程中, 熵编码是最后一道操作, 其压缩能力大约为 2~3 倍, 可见, 对具有 25 倍压缩率的 JPEG 图像来说, 熵解码后还有 10 倍左右的压缩率, 其数据量与非压缩格式的数据量相比, 仍然是较小的。

(3) 反量化后：有损编码时, 一般都有量化步骤, 因此在进行图像的高精度处理时, 除了要进行熵解码外, 还需要进行反量化, 以提高图像处理的效果, 在反量化后进行压缩域的图像处理也是常用的方法。

图 3.12 是一个典型的编解码过程示意图<sup>[20]</sup>。位置 0 为像素域特征提取操作位置, 可以看出, 对于压缩图像, 若在位置 0 提取特征, 首先需要进行全解码。在压缩域操作中, 所谓不解码, 就是在位置 1 未进行熵解码之前进行压缩域的处理; 而少量解码, 就是在位置 2、3 进行熵解码或反量化之后进行处理。即使在位置 2 或位置 3 进行, 计算量也可大大降低, 因反变换所需时间一般占总解压时间的 40%~60%。

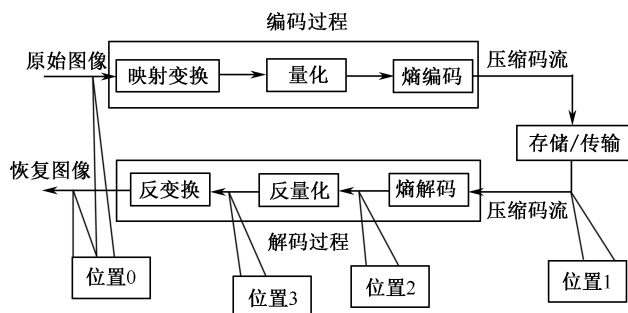


图 3.12 一个典型的编解码过程示意图

### 3.1.4 压缩域图像检索的研究内容

---

对已有压缩算法所形成的压缩数据的检索技术已经进行了许多年,也取得了一些成果,但由于现在的压缩算法在最初的设计实现时并未考虑后续图像检索处理的要求,所以在这些压缩数据上实现图像检索的能力是十分有限的,一个根本的解决方法是从图像压缩的角度进行压缩域图像检索技术的研究。从本质上讲,图像压缩编码算法的主要目的是寻求图像数据的有效表征形式,这种表征不仅要具有小的存储要求,同时还能较好地保持原图像数据的信息含量。而从支持图像检索的角度来看,由于不同压缩算法所形成的压缩数据中包含了表征原始图像主要内容的大部分信息,所以理论上任何压缩码流都能对压缩域图像检索提供支持,而其区别主要在于压缩域图像检索处理过程实现的复杂度不同,且不同压缩算法对不同的图像检索操作具有不同的支持程度。因此,支持图像检索的压缩算法的研究就是要选择和实现有效的图像信号表征形式和量化方法,使其不仅有利于高效压缩的实现,同时还能够从其所形成的压缩数据(不解码或尽量少解码)中得到或构建出图像内容的特征量用于支持图像检索操作。总体来说,基于压缩域的图像检索技术的研究内容主要有两个方面<sup>[22]</sup>。

(1) 从处理角度看,由于现有的编码算法没有考虑到压缩域的分析处理,因此必须深入分析压缩域的特有性质,研究其压缩域究竟有什么处理能力,力争在不解码或不完全解码的情况下进行处理,以提取特征信息进行图像检索。由于目前存在大量压缩数据,这方面的研究具有非常现实的意义。传统的提取特征信息的处理技术比较成熟,一个很自然的想法就是如何在压缩域推导出与这些处理算法相对应的对等操作<sup>[23]</sup>。例如,对于传统的空域处理中的一些技术就很容易推导出在以 DCT 为基础的压缩域中的对等操作,只是把它变成频率域处理而已。但并不是每一种传统方法都能找到对应的压缩域模式,也不是压缩域的处理模式全是传统模式对应的翻版,人们还要根据压缩域的特点,研究更符合实际应用的压缩域特有的图像处理新算法。

(2) 从编码角度看,研究新一代的编码算法,使其既具有较高的压缩率,又具有支持压缩域检索与处理的能力,即同时考虑编码与处理,在更高的层次上来解决图像/视频信息的管理和操纵,这将是新一代编码的研究方向,具有前瞻性。

### 3.1.5 压缩域图像检索的研究方法

---

基于压缩域的图像检索技术是多媒体领域的一个非常活跃的研究方向,但是由于需要在压缩域直接进行分析处理,因此受到压缩算法本身局限的影响,这给研究工作带来了相当大的难度。针对压缩域的图像检索技术的研究,可归结为寻求传统处理方法在压缩域的对等操作与寻求压缩域的特有操作两种方法。



## 1. 寻求对等操作

这里所说的对等操作,指的是能够通过严格的数学推导,将一个域的操作转换到另一个域的操作。由于传统的图像处理方法的处理对象是基于原始像素域的,因此,许多空域处理操作很容易在频率压缩域中找到对应的处理方法,如基本的标量加、标量乘、向量加、向量乘运算等。许多压缩域的对等操作,都可通过严格的推导来证明,但在压缩域寻求空域的对等操作的过程中,一定要看其算法的简易性、有效性和对计算资源的需求,其综合指标(主要指处理时间、处理效果、耗费资源等情况)至少要好于或近似等于“解码”+“传统处理技术”+“再编码”的传统处理模式,否则没有实际意义,这是研究基于压缩域的图像检索技术的根本出发点。

## 2. 寻求特有操作

并不是所有的处理算法在压缩域中都具有对等操作,因此,如何根据压缩域所具有的特点寻求压缩域所特有的图像处理算法是另一个研究途径。对已有的压缩算法,深入分析压缩域所具有的特点,设计出方便、实用、高效的处理方法;在设计新的支持压缩域处理的图像编码算法时,充分考虑到压缩域的特性,提供具有更多操作功能的压缩码流。寻求压缩域的特有操作是压缩域检索技术中研究的难点,也是重点。

基于压缩域的图像检索方法与原始图像所采用的压缩方法有密切关系,用什么方法压缩对其后的检索有重要影响。考虑到常用的压缩技术的特点和压缩域检索的要求,可将压缩域检索技术分为空间压缩域技术、变换压缩域技术和融合压缩域技术三大类,如图3.13所示<sup>[24]</sup>。空间域方法主要包括矢量量化、分形编码和预测编码,变换域方法主要是利用DFT、DCT、小波变换和K-L变换等。目前大多数的研究成果都集中在变换域上,尤其是基于小波压缩域和DCT压缩域。

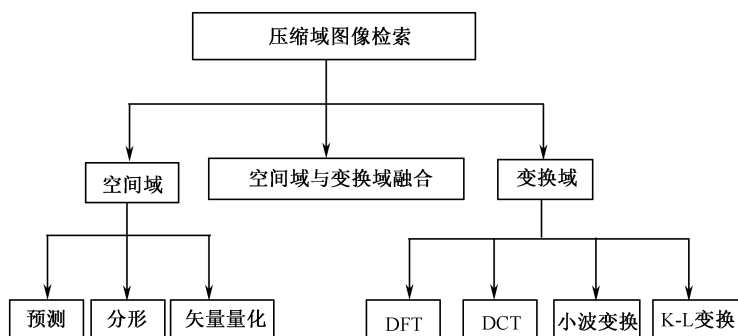


图 3.13 压缩域图像检索技术

## 3.2 空间压缩域技术

### 3.2.1 矢量量化

矢量量化是图像编码领域一个重要的编码方法。其中心思想是通过训练集合首先构造一个有限集看作码书，然后对任意输入图像中的每一分块矢量按一定的距离准则进行分类，并且将原图像中的所有分块矢量都采用与之相匹配的特征矢量码字来进行表征，而与所有分块矢量相匹配码字的索引序列就是最终的压缩数据。这样的编码结构和压缩数据非常适合与图像检索技术相结合。从结构上说，矢量量化本质上是一个聚类和分类的过程，这样的结构天然地和图像聚类分析及分类处理等操作的要求相一致。从获得的压缩数据看，由于矢量量化是通过一个有限数量的再生矢量集合来对所有输入图像进行表征的，它的码字在一定程度上与图像的内容相对应，该矢量集合也描述了解压缩图像的性质，所以对该矢量集合（码书）的任何分析处理操作都可以间接地反映到解压缩图像中。正是因为矢量量化具有的这种良好特性非常有利于图像的检索和查询，使得其可以广泛应用于图像增强、图像分类、边缘检测及图像检索等方面<sup>[25]</sup>。

矢量量化的理论基础是香农的速率-失真理论。基本的矢量量化器可以定义为从  $d$  维欧氏空间  $\mathbf{R}^d$  到其一个有限子集的一个映射，即  $Q: \mathbf{R}^d \rightarrow \mathbf{C}$ ，其中， $\mathbf{C} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N | \mathbf{y}_i \in \mathbf{R}^d\}$  称为码书， $N$  为码书大小。该映射满足  $Q(\mathbf{x} | \mathbf{x} \in \mathbf{R}^d) = \mathbf{y}_s$ ，其中， $\mathbf{x} = (x_1, x_2, \dots, x_d)$  为  $\mathbf{R}^d$  中的  $d$  维矢量， $\mathbf{y}_s = (y_{s1}, y_{s2}, \dots, y_{sd})$  为码书  $\mathbf{C}$  中的码字并满足

$$\text{dist}(\mathbf{x}, \mathbf{y}_s) = \min_{1 \leq j \leq N} \text{dist}(\mathbf{x}, \mathbf{y}_j) \quad (3-5)$$

其中， $\text{dist}(\mathbf{x}, \mathbf{y}_j)$  是输入矢量  $\mathbf{x}$  与码字  $\mathbf{y}_j$  之间的失真测度（距离）。每一个矢量  $\mathbf{x}$  都能在码书  $\mathbf{C} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$  中找到其最近码字  $\mathbf{y}_s = Q(\mathbf{x} | \mathbf{x} \in \mathbf{R}^d)$ 。输入矢量空间通过量化器  $Q$  量化后，可以用划分  $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N\}$  来描述，其中， $\mathbf{S}_i$  是所有映射成码字  $\mathbf{y}_i$  的输入矢量集合，即  $\mathbf{S}_i = \{\mathbf{x} | Q(\mathbf{x}) = \mathbf{y}_i\}$ 。这  $N$  个子空间  $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N$  满足

$$\bigcup_{i=1}^N \mathbf{S}_i = \mathbf{S} \text{ 且 } \mathbf{S}_i \cap \mathbf{S}_j = \Phi (i \neq j) \quad (3-6)$$

基本的矢量量化编码和解码过程如图 3.14 所示。矢量量化器根据一定的失真测度（距离）在码书中搜索出与输入矢量之间失真最小的码字，在二次存储设备中存储时仅存储该码字的索引。矢量量化解码过程很简单，只要根据码字索引在码书中查找该码字，并将它作为重构矢量。内存容量能够容纳下码书时，码书一般存储在内存中。矢量量化之所以能够压缩数据，是由于它能够去除数据之间的冗余度，有效

地利用矢量中各分量之间相互有联系的4种性质：线性依赖、非线性依赖、概率密度函数的形状和矢量维数的本身。

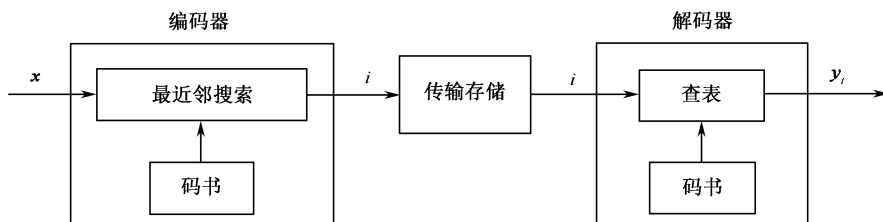


图 3.14 基本的矢量量化编码和解码过程

从信源的角度来看，矢量量化的结果可以认为是一个  $N-1$  阶有记忆信源所发出的一个符号集合，每个符号是一个具有统计关联特性的  $N$  维矢量，对相似的图像所对应的符号集合来说，同一符号的概率分布应接近。在矢量量化过程中，任一输入图像的分块矢量都依据统一的距离准则映射到有限特征矢量集合（公共码字）的某一码字上，所以不同的图像与该有限矢量集合之间映射的统计特性在一定程度上反映了图像的内容。Idris<sup>[26]</sup> 基于这样的思想，使用频数统计直方图构造了图像特征矢量集合中的矢量，并将其作为表征图像内容的特征量，用于比较查询图像和图像库中目标图像的相似度，从而实现了将矢量量化压缩数据直接用于支持图像检索的目的。同时，为了进一步降低对直方图特征量的存储要求及直方图相互间距离计算的复杂度，还对码字使用频数直方图进行了简化处理，提出使用图特征量来对码字进行简化处理的方法<sup>[27]</sup>，即某一码字矢量，如果在图像压缩过程中使用过，则与该码字对应的特征位置为 1，否则为 0，而不同图像间相似程度的计算只需对其使用不同的图特征量按位进行异或操作，即可大大减小特征矢量的存储量和距离计算的复杂度，但这样简化处理的代价是降低了图像检索的正确率。图 3.15 和图 3.16 分别给出了上述两种统计特征量的生成示意图<sup>[13]</sup>。

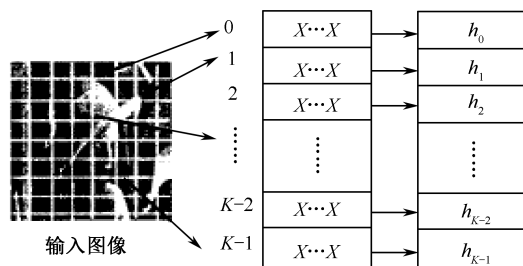


图 3.15 基于矢量量化的码字使用频数直方图特征构造示意图

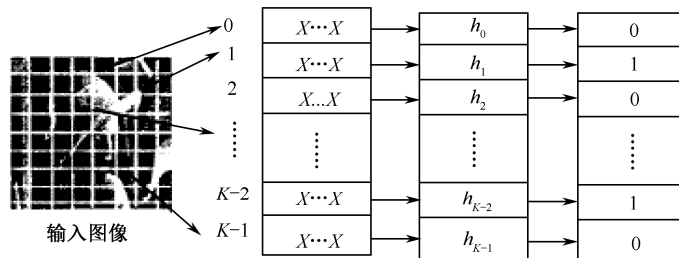


图 3.16 基于矢量量化的码字使用图特征量的构造示意图

文献[28]探讨了大型雷达图像数据库的有效表征问题，目的是为了优化存储量和搜索时间，利用树状结构矢量量化的聚类方式，并用树状结构矢量量化下的多分辨率小波来表征图像特征，这种树状结构是根据脉冲的多分辨率小波分解结果引入的，因此该方法的整体性能较好。文献[29]将矢量量化技术应用到基于内容的遥感图像检索中，采用估测多种形变的方式来提高作为内容描述符 VQ 码字的性能，通过测试两种查询（按类查询、按数值查询）的性能，发现后者的性能更为卓越。魏海等<sup>[30]</sup>将矢量集中进行分类，对不同类别的矢量进行不同的处理，既具有较高的压缩率，又能进行压缩域的图像检索。文献[31]设计了一种利用矢量的索引值提取图像连续区域直方图进行图像检索的方法，具有较好的旋转、平移和尺度不变性。文献[32]结合投影方法与矢量量化技术提出一种基于局部投影与块 LBP 特征的图像检索方法，首先对图像库中所有图像进行子块的划分，并对每一子块进行投影，得到投影矢量训练集合  $\{P_i | 1 \leq i \leq D\}$  ( $D$  为投影矢量总数)，对投影矢量训练集合  $\{P_i | 1 \leq i \leq D\}$  按照 LBG 算法生成码书  $C_N$  ( $N$  为码书大小，一般为 2 的幂次方)，对输入查询图像，划分子块并对每一个子块进行投影得到投影矢量，然后在码书  $C_N$  中得到最匹配的码字，统计所有图像子块的匹配码字所对应的索引序号，最后生成投影矢量的索引直方图  $H(v_1, v_2, \dots, v_i, \dots, v_N)$  ( $v_i$  表示第  $i$  个码字为最匹配码字的频数， $N$  为码书大小，即码书中所有码字的总数)。提取的投影矢量索引直方图特征能有效提取图像的颜色分布和空间分布等信息。

若根据图像中矢量的特性，先进行矢量分类，例如，根据矢量方差值，把矢量分成平滑矢量（方差较小）和细节矢量（方差较大），不同类矢量对应不同的码本，矢量量化时根据矢量的类型到相应的码本中寻找其码字索引，这种矢量量化称为分类矢量量化。作为一种特殊的矢量量化方式，分类矢量量化是由 Ramamurthi B. 所提出的<sup>[33]</sup>，其主要目的是为了克服在高压压缩率下矢量量化所出现的边缘退化失真现象。主要思想是首先对图像的分块矢量集合按一定的特征进行分类，不同类别的矢量块子集根据其重构图像质量的重要程度分别采用不同长度的类别码书进行矢量量化，这样可以保证那些对人眼观察较重要的类别矢量块能够获得足够高的量化精度，从而在总体上能够获得主观质量更好的解码图像。基于分类矢量量化的检索方法与上述基于矢量量化

的方法类似,不同之处仅在于分别统计各类码本中码字矢量使用频数的统计直方图,然后把它们连成一个统一的直方图。基于分类矢量量化进行检索有一个优点,可以根据人眼对各类矢量的敏感程度,对直方图中的分量采用不同的权值来提高检索效率。魏海等<sup>[30]</sup>比较了标量量化(SQ)、矢量量化(VQ)和分类矢量量化(CVQ)用于检索的性能,结果表明,VQ的检索性能优于SQ,而CVQ的检索性能明显优于VQ。

### 3.2.2 分形编码

分形的理论基础是法国数学家 Mandelbrot 所创立的分形几何学<sup>[34]</sup>,这里主要论述分形编码的思想。对于二维的灰度图像,其分形压缩变换可定义为

$$W_i(R_i) = s(R_i)S^{(n+1)}\tau(R_i)D(R_i) + o(R_i) \quad (3-7)$$

其中,  $D(R_i) \in \{D_1, D_2, \dots, D_m\}$  是域块;  $\tau(R_i) \in \{\tau_1, \tau_2, \dots, \tau_8\}$  是 8 种对称变换;  $S^{(n+1)}: \mathbf{R}^{2^{2(n+1)}} \rightarrow \mathbf{R}^{2^{2n}}$  是下抽样算子;  $s(R_i) \in \{s_1, s_2, \dots, s_n\} \subset \mathbf{R}$  是缩放系数;  $o(R_i) \in \{o_1, o_2, \dots, o_l\} \subset \mathbf{R}$  是偏移量;  $R_i$  是值块。

由以上定义可以看出,分形是一类无规则、混乱且复杂,但其局部与整体具有相似性的系统,所以分形最主要的特征就是其所具有的自相似性。这种自相似性既可以是形态或结构几何意义上的相似,也可以是信息或功能统一意义下的相似。而在实际中,具有统计意义下的自相似分形集合占大多数<sup>[35]</sup>。除了具有自相似性这一特性外,分形集合所具有的另一个诱人之处在于对大多数的分形集合来说,尽管表面很复杂,但都能够从少数参数出发按极简单的规则来生成它,也正是这一点促使人们将分形应用于图像的压缩编码中。因为所有的图像都可以认为是具有自相似性的分形系统,如果我们能够对一幅图像找到它所对应的生成规则和参数,那么这些参数和规则便唯一地表示了这幅图像,而这些参数和规则所具有的数据量相对于原始图像的数据量而言是极其微小的,这样就能够获得极高的压缩率。当然,严格来说,我们不可能绝对得到任一幅图像所对应的分形规则,但至少统计意义下我们可以选择一些规则得到一幅图像的分形近似图像来足够好地接近于原始图像。正是由于这些原因,分形理论一经引入图像压缩领域便引起了众多研究人员的普遍关注,相继出现了许多令人瞩目的分形编码方法。

另外还有一种迭代分形编码,其理论基础是分形几何学中的迭代函数系统理论,迭代函数系统(IFS)通常是指在某一度量空间内的收缩仿射变换集  $W = \{w_i | i = 1, 2, \dots, N\}$ ,它是一种通过寻找信号自身递归的变换不变关系来构造自相似分形的方法。根据 Banach 收缩映射不动点定理,只要存在常数  $0 < s < 1$ ,使得

$$d[w_i(x), w_i(y)] < sd(x, y), \forall i, x, y \quad (3-8)$$

就存在唯一的吸引子集  $A$ , 定义为

$$A = W(A) \Delta \bigcup_i w_i(A) \quad (3-9)$$

$A$  满足自相似性和吸引性: 任何集合  $B$  经过收缩仿射变换算子集  $W$  的反复作用都会收敛到  $A$ ; 同时, 不论从  $B$  中的哪一点开始, 只要按同样的顺序嵌套 IFS 中的变换, 都会收敛到  $A$  中的同一点。

分形编码算法主要利用了图像内部的自相似性, 具有很高的压缩率, 但由于编解码巨大的不对称性, 使其应用受到了限制。在基于块的分形编码算法中, 输入图像被分割成互不重叠的 Range 块, 对每个 Range 块, 选择一个 Domain 块, 这个 Domain 块能通过一些仿射变换来最大限度地相似对应的 Range 块, 这些仿射变换就是分形码, 它以高压缩的形式表征了这幅图像。可见, 分形编码实际上是一种自我索引技术, 基于分形编码的索引算法的研究主要就是利用了这一特点。

Sloan<sup>[36]</sup>首先发现了分形图像编码技术在基于内容的图像检索中的应用前景, 并进行了探索, 提出了将图标图像数据库中的图像直接拼凑起来进行编码的方法, 即对于任何图像中的排列块可以在任何图像中找到区域块, 而两幅图像之间的相似性则可以用一幅图像中的子块当作另一幅图像中区域块的数目来度量。对这方面的进一步研究是由 Zhang 等人开展的, Zhang 等人<sup>[37]</sup>提出直接利用图像的分形码来实现基于纹理的图像检索技术, 这里分形码是以图像子块为单位得到的, 即首先把图像分成不重叠的若干子块, 对每一子块根据分形原理寻找其分形码, 检索时直接比较查询图像和目标图像的分形码。文献[38]提出了一种基于九分树分解和分形编码的图像检索策略, 并进一步比较了基于分形的方法与小波子带能量法的检索效果, 结果表明, 对于纹理图像, 小波方法要好些, 而对于其他图像, 则分形方法能略胜一筹。Ida<sup>[39]</sup>还提出了一种基于分形码的图像分割方法, 主要是基于以下设想: 如果  $S$  区域中有 Domain 块, 则其相应的 Range 块也在这个区域内, 因此可以由分形码分割出区域  $S$ , 这完全可用到图像检索技术中。王志勇等人<sup>[40]</sup>提出了一种基于块限制的分形编码算法和匹配策略, 并将它们用于图像检索, 首先图像被预先分成互相不重叠的子图像块, 然后对这些子图像块进行独立的分形编码, 从而获得整幅图像的分形码, 采用改进的基于九叉树的分配策略进行图像间相似性的匹配, 避免全局地进行分形码的匹配, 实验结果表明, 该算法计算量小, 在计算时间和存储时间上都有较高的效率。

文献[41]利用分形对图像所具有的独特表征能力, 对基于迭代分形的压缩方法和检索方法进行了研究, 在小波变换域内提出了一种基于迭代分形的高效图像压缩方法, 并在图像分形码的基础上, 利用迭代函数系统的分布特性来进行图像间相似距离的计算, 从而实现了基于迭代分形压缩数据的图像检索操作。文献[42]利用图像的分形编码数据里包含的空间域信息, 提出了一种在分形域内基于内容的图像检索方法, 该方法在分形编码的基础上定义了提取图像与偏移量图像, 通过计算两者的距离来求原始图像和目标图像的相似度。文献[43]提出了一种利用分形编码重要的拓扑特性来处理图像索引的新方法, 即先将图像经分形编码得到图像的迭代函数, 然后将其编码

图像存入数据库中,成为该图像的索引文件,最后通过对此索引文件的比对来找出与查询图像相似的图像,实验显示,图像经过分形编码所表现出的几何性质及独特的有效性和鲁棒性证明该方法是一个更有效率、准确度高的检索方法。文献[44]提出了一种图像熵和分形编码相结合的图像检索方法,首先计算图像熵和比较设定的阈值对图像库进行预分类,然后利用 Jacquin 方法计算得到查询图像的分形 IFS 编码,把图像库同类图像作为初始图像进行分形迭代解码,最后计算解码图像与查询图像的相似距离得到检索结果,实验结果表明,该方法在基本保证图像检索效率的前提下,极大地提高了检索速度,具有很好的有效性和可行性。

### 3.2.3 预测编码

预测编码的基本思想是依次扫描图像中各个像素,提取每个像素中的新信息,对这些信息编码来消除像素中的冗余。这里像素的新信息定义为该像素的当前值与预测值的差。

预测编码可分为无损预测编码和有损预测编码。无损预测编码包括两个步骤:预测和编码。预测时,逐个读取输入图像像素序列  $f_n (n=1,2,\dots)$  中的每个像素,预测器根据若干个过去的输入产生当前输入像素的估计值。例如,可通过将  $m$  个先前的像素进行线性组合以得到预测序列,即

$$\hat{f}_n = \text{round} \left( \sum_{i=1}^m a_i f_{n-i} \right), \quad n=1,2,\dots \quad (3-10)$$

其中,  $m$  是线性预测器的阶;  $\text{round}$  是舍入函数;  $a_i$  是预测系数。由于图像内像素的相关性,预测序列与实际序列的差(即预测误差)序列  $e_n (n=1,2,\dots)$  的概率密度函数一般在零点有一个高峰。对预测误差序列再用符号统计编码的方法编码可以得到压缩的结果。

一般对预测编码图像的检索方法有两种:一种仅将编码结果解码(部分解码但不变动预测结果),另一种直接利用编码结果(完全不解码)。

(1) 仅将编码结果解码的方法。如果仅将编码结果解码,得到的是预测误差序列。对每个像素来说,对它的预测误差是对它邻域的一种统计描述,即该像素相对其邻域的变化量。因为图像中邻域的性质与纹理有密切关系,所以可以认为预测误差在一定程度上反映了纹理信息,可以用来进行基于内容的图像检索。

(2) 直接利用编码结果的方法。在无损预测编码的第二个步骤中常用统计(符号)编码方法,如霍夫曼方法等,统计编码的特点是把较短的码赋给出现概率较大的预测误差值,而把较长的码赋给出现概率较小的预测误差值。具体来说,一个霍夫曼码本身就包括每个预测误差值所赋的码字,而每个码字的长度与它所赋予的预测误差值的出现概率成反比。所以,霍夫曼码也基本包含了预测误差值直方图中的信息,可以用

来进行基于内容的图像检索。

## 3.3 变换压缩域技术

### 3.3.1 基于 DFT 压缩域

DFT 是图像处理和信号分析处理中的一个非常重要的数学变换, 由于 DFT 利用复数指数函数进行, 其能量压缩特性良好, 从而可提供良好的编码性能。DFT 的某些性质在检索和模式匹配中是很有用的, 如系数幅值的平移不变性、简化空域的相关运算等。

DFT 正反变换的定义为

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}nk}, \quad x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}nk} \quad (3-11)$$

Stone<sup>[45]</sup>等人提出了一种在 DFT 压缩域的图像检索算法, 这种算法合理组织傅里叶系数; 同时, 提供两个阈值, 用户可通过这两个阈值来独立地调整匹配的近似度, 使用一个阈值来控制亮度匹配, 再用一个阈值来控制纹理的相似程度, 用这两个阈值可以使匹配达到由粗到精的过程。另外, 由于这两个阈值之间是相关的, 故利用傅里叶系数可以有效地计算出图像的特征, 且在傅里叶系数大部分是零的时候, 这种计算尤其有效。文献[46]利用 DFT 系数的统计特性和 DFT 系数径向分布结合角度分布实现了卫星图像的纹理分类; 同时, 还研究了基于傅里叶系数半径(即幅度)和相位分布特性的检索性能。DFT 系数的径向分布在一定程度上反映了图像纹理的粗糙程度, 而 DFT 系数的角度分布则反映了图像纹理的方向性。另外, 文中还对比了两种方法的纹理分类性能, 结果表明, 当图像中存在明显突出的频率成分时, 基于 DFT 系数径向和角度分布的方法具有较好的分类效果; 而当频率成分比较均匀时, 基于统计的方法则具有更好的分类效果。文献[47]提出了一种基于 DFT 和 log 极坐标变换的描述子, 该描述子不仅具有平移、旋转和尺度不变性, 而且实现了对图像内容的综合概括, 包括图像的形状、纹理和颜色内容, 从而克服了上述的缺陷, 得到了优良的检索结果。文献[48]提出了一种基于多尺度拱高形状描述的图像检索方法, 多尺度拱高示例如图 3.17 所示<sup>[48]</sup>。该方法用多尺度拱高来度量形状轮廓线在其每一个点的弯曲程度, 在每一个尺度级, 为形状轮廓线构建一个一维的拱高函数, 然后对其进行傅里叶变换, 组合各尺度级的低阶傅里叶变换系数, 构成描述形状的特征向量, 该方法不仅能描述形状的全局特征, 而且能描述形状的细节信息, 因而在图像检索应用中具有较高的检索效率。



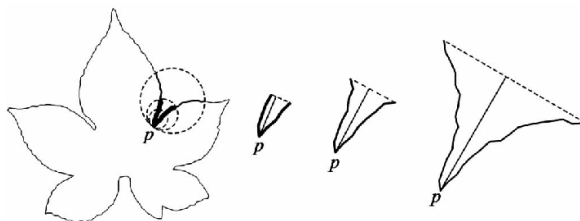


图 3.17 多尺度拱高示例

使用极坐标下的傅里叶系数分布来进行纹理图像的分类也是常用的算法之一。文献[49]研究了基于 DFT 系数角度分布的图像检索方法,先对图像进行低通滤波的预处理;然后进行傅里叶变换;在  $180^\circ$  范围内扫描 DFT 频谱图,统计每一角度范围内的 DFT 系数之和,得到一个角度分布直方图,以此作为特征矢量进行图像检索。虽然该特征矢量具有平移不变性,但是在图像发生旋转后,它将发生循环移动,因此该方法不具有旋转不变性,对这一缺点可通过对特征矢量进一步作 FFT 使其具有旋转不变性。

虽然基于 DFT 压缩域的检索技术取得了较好的结果,但是由于 DFT 是一个复数变换,运算量大,实现困难,另外对于一般的图像信号,利用 DFT 进行压缩,效果也不是很理想,因此很少被用来进行图像压缩,这也是图像压缩标准中没有采用 DFT 的原因。从而,基于 DFT 压缩域的图像检索能力也就无法很好地体现,目前仍有学者在作这一方面的研究,但是只是利用 DFT 作为特征提取的一种手段,不再强调压缩域检索这一概念<sup>[6]</sup>。

### 3.3.2 基于 DCT 压缩域

#### 1. 概述

DCT 是从 DFT 演变而来的,对许多自然图像来说,DCT 接近最优的 K-L 变换,是现在应用最广泛的多媒体数字压缩技术,图像、视频压缩国际标准 JPEG、MPEG 和 H.261 都采用了 DCT 技术。DCT 有许多优点。首先,它是一种线性正交变换,变换核各矢量间单位正交,可以将  $8 \times 8$  图像的空间域转换为频率域,只需要用少量的数据点表示图像;其次,二维 DCT 是对称分离的,即二维 DCT 可分解为行列方向的两维一维 DCT, DCT 有快速算法,算法性能很好,易于在软硬件中实现,而且 DCT 算法对称,利用逆 DCT 算法可以解压缩图像;最后, DCT 产生的系数很容易被量化,因而能获得好的块压缩,性能优于 DFT 等其他变换,去相关压缩能力接近于 K-L 最佳变换。更为重要的原因是,经过 DCT 后所得 DC 系数和 AC 系数分别具有以下特性。

- (1) 对于每个变换块,每一个分量[直流(DC)分量]表达了此块的平均亮度信息。
- (2) 低频信息集中于变换块的左上角,高频分量位于变换块 Zigzag 排序的后方。
- (3) 交流(AC)系数反映了像素间差异的频率信息和方向信息。

基于 DCT 压缩域的检索技术就是在现有的压缩标准的基础上,通过分析 DCT 系

数及压缩算法特点, 力争在不解码或部分解码的情况下实现图像检索。利用 DCT 良好的数学变换性质能够在 DCT 域进行高效的图像检索, 避免了传统空域操作烦琐的解 DCT 压缩数据, 处理后再还原为 DCT 压缩数据的操作过程, 将其应用于图像检索能够充分发挥检索准确率高、检索速度快、计算复杂度低、计算量小等优势。

在 DCT 压缩域进行图像检索, 主要是分析、提取能反映图像内容的 DCT 系数的特性。目前大致有两类方法<sup>[50]</sup>: 第一类方法是对 DC 图进行处理, 因为 DC 图是整幅图像的缩略图, 表达了图像的平均能量, 所以可采用传统的图像检索方法(如灰度直方图等), 但是没有利用 DCT 域的低中频信息; 第二类方法是根据交流(AC)系数反映了像素间差异的频率信息和方向信息, 通过计算 AC 系数的能量直方图等信息来检索图像, 其检索效果并不太理想。应该说目前对基于 DCT 压缩域的图像检索技术研究还处于比较初级的阶段, 还只是对 DCT 数据的简单直观应用。例如, 根据 DCT 系数的 3 个大致方向性分布特征, 初步判定边缘的存在与否和边缘方向, 用直流系数的统计直方图作为图像相似性测度和判据等。

## 2. 纹理特征的提取

基于 DCT 压缩域的检索技术中研究最多的是纹理特征的提取及基于纹理特征的检索和分类。Smith<sup>[51]</sup>首先提出一种基于 DCT 的纹理图像检索方法, 它将图像分成  $4 \times 4$  的子块, 对其进行 DCT, 对所有块在同一位置上的系数计算其均值和方差, 构成一个具有 32 个分量的特征矢量, 用它来表征整个纹理的图像特征, 并用 FDA (Fisher Discriminate Analysis) 来进行降维处理。由于这种方法采用的 DCT 块大小和压缩标准不兼容, 要使用这种方法时, 必须完全解码才能提取特征, 但它给我们提供了一种 DCT 压缩域图像检索的思路。Reeves<sup>[52]</sup>在此基础上进行改进, 采用与标准兼容的  $8 \times 8$  子块, 并且认为反映不同纹理特性的主要是前几个 AC 系数, 从而避免了判别分析, 它只对每个块内的前 8 个系数计算方差来形成特征矢量, 由于这种方法生成的特征矢量维数比上一种方法少, 所以它的算法时间复杂度也比上一种方法小。与上述计算单个通道均值和方差的策略不同, 文献[53]根据 DCT 系数的特性, 把 DCT 系数分成不同的区域, 通过统计这些区域的能量得到一个具有 9 个分量的综合特征。Sim 等人<sup>[54]</sup>引入了人眼视觉特性, 利用一个符合人眼视觉掩模对各个频率通道的特征分量作加权处理。文献[55]等考虑到直方图技术在图像检索中的作用, 提出了一种基于低频 DCT 系数能量直方图的检索方法, 该方法首先选择 DCT 块中的一个对称区域, 然后统计该区域中的 DCT 系数的能量直方图, 并以此作为索引进行图像检索。

上述各种方法所提取的纹理特征抗干扰能力通常不是很强, 在图像压缩倍数较高或有噪声的情况下, 检索及分类效果往往不好。为此, 黄祥林<sup>[56]</sup>适当组织 DCT 系数使之反映图像纹理的方向性, 提出了一种基于 DCT 区域能量方向性的纹理图像分类方法。首先根据一定区域的变换系数代表着一定方向的频谱成分, 将  $8 \times 8$  子块 DCT 系数分成直流区、竖直纹理区、水平纹理区和对角纹理区, 如图 3.18 所示。4 个区域

的能量直接从 RLE 码流中采用下式来计算。

$$\left. \begin{aligned} E_0 &= \sum_{m=1}^N (A_{00}^m - \bar{A}_0)^2 \\ E_k &= \sum_{m=1}^N \sum_{\Omega_k} (A_{ij}^m - \bar{A}_k)^2, \quad k=1,2,3 \end{aligned} \right\} \quad (3-12)$$

其中,  $\bar{A}_0 = \frac{1}{N} \sum_{m=1}^N A_{00}^m$ ;  $\bar{A}_k = \frac{1}{21N} \sum_{m=1}^N \sum_{\Omega_k} A_{ij}^m$ ;  $\Omega_k$  ( $k=0,1,2,3$ ) 分别表示原图像中直流成分区域、水平方向的区域、对角方向的区域和垂直方向的区域, 并且  $\Omega_k$  中的系数个数为 21。

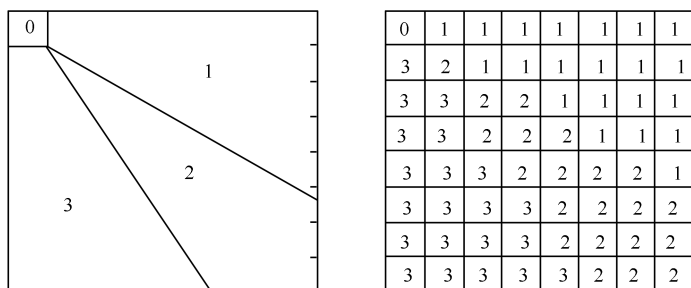


图 3.18 DCT 域的能量分区

最后通过一种多分辨率特征组织形成特征矢量来支持纹理分类及检索。实验结果表明, 这种方法不仅计算简单, 而且在有噪声的情况下也具有很好的检索结果。

黄祥林还提出一种基于 DCT 频率系数分布特点的纹理分类方法<sup>[57]</sup>, 对  $8 \times 8$  子块 DCT 系数按低频率、中频率和高频率分区, 从 RLE 码流中直接统计这些区域的能量, 形成多分辨率特征矢量。由于这里是把各个方向中同等地位的系数考虑在同一个区域内, 因此这种方法具有较好的旋转不变性。

这两种方法的另外一个特点是结合 JPEG 压缩算法的具体特点, 直接从 RLE 码流中计算特征矢量, 真正实现了基于压缩域的图像检索。

Fan<sup>[58]</sup>等直接在 DCT 域提取有方向性的图像特征进行检索, 并指出 DCT 域图像特征信息还需要进一步的开发和利用。Feng 等利用归一化的区域面积向量进行 JPEG 图像的检索<sup>[59]</sup>。

Climer 等<sup>[60]</sup>利用 DCT 系数中的 DC 系数和少量重要的 AC 系数形成图像的四叉树索引结构 (如图 3.19 所示), 然后采用逐层相似性比较的方法, 加快了检索速度。

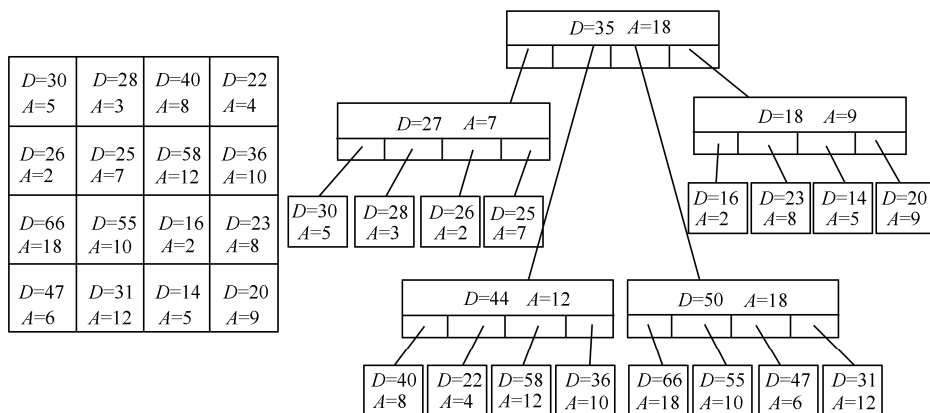


图 3.19 四叉树构造图

图 3.19 中的四叉树构造方法只利用了块中的 DC 系数和块中最大的 AC 系数，4 个相邻的块分别作为子节点，又构造出一个父节点，父节点中的  $D$  值是 4 个子节点的  $D$  值的取整均值， $A$  值则是子节点中  $A$  值的最大值。相邻的父节点又类似地构造出祖父节点，直到将所有的块全部组合成一棵四叉树，这种方法具体的距离计算公式如下。

$$D = \sum \left[ (D_q - D_n) + \left( \frac{A_q - A_n}{\lambda} \right) \right] \quad (3-13)$$

其中， $D_q$  和  $D_n$  分别是被查询图像和查询图像对应节点的平均 DC 值； $A_q$  和  $A_n$  分别是相应的最大 AC 系数。相似性比较时，从根节点开始逐层比较，对应节点的累加值小于预设的门限值，就认为是相似的。当图像尺寸很大时，四叉树也很大，当将全部节点比较完时，很费时间，所以不一定要将所有的节点全部比较完，否则运算量很大；同时，四叉树的存储会占用很大的空间，这种方法对图像亮度的改变比较敏感。

Jose 等<sup>[61]</sup>利用 DCT 系数的量化值去构造能量直方图，提出一种 DCT 系数量化直方图算法。量化后的 DCT 系数是一个整数，其最大值是直方图的维数长度。一个  $8 \times 8$  块的能量直方图表示如下。

$$H_c(m) = \sum_u \sum_v \begin{cases} 1, & F^Q(u,v) = m \\ 0, & \text{其他} \end{cases} \quad (3-14)$$

其中， $F^Q(u,v)$  表示 DCT 系数在  $(u,v)$  的量化值。

量化系数能量直方图体现了不同的能量强度在所有的块中所发生的次数。不同图像有着不同的能量强度水平，相似的图像有可能在能量强度层次上相似，Jose 的这种方法就利用了这一点。它的相似度衡量方法为归一化的  $L_1$  距离方法，即

$$D(H_q, H_c) = \frac{1}{n} \sum_{i=1}^n \frac{|h_i^q - h_i^c|}{M} \quad (3-15)$$

其中， $H_q$  和  $H_c$  分别是被查询图像和查询图像的量化系数能量直方图； $n$  为直方图的

维数;  $M$  是在某个能量强度上系数发生的最大次数。这种方法不需要反量化,可直接利用量化的 DCT 系数作统计,比四叉树方法要简捷得多,不过事先要确定直方图的长度,即最大能量强度。

文献[62]提出一种基于 DCT 和 SVD 的图像检索算法,首先将图像分割为  $n$  个互不相交的子块,利用离散余弦变换提取重要系数作为子块颜色特征,进而对图像进行如图 3.20 所示的区域划分,将每个区域中的子块颜色特征分量组成矩阵进行奇异值分解,得到该区域的检索特征向量,从而完成图像检索。文献[63]提出一种在压缩域中基于随机变量数字特征累加直方图的图像检索算法,特征提取过程如图 3.21 所示,通过离散余弦变换将图像变换到变换域,计算图像块的数学期望和方差,建立累加直方图,利用相关性判定两幅图像的差异。实验结果显示,此种方法检索精度高、速度快、计算量较小。

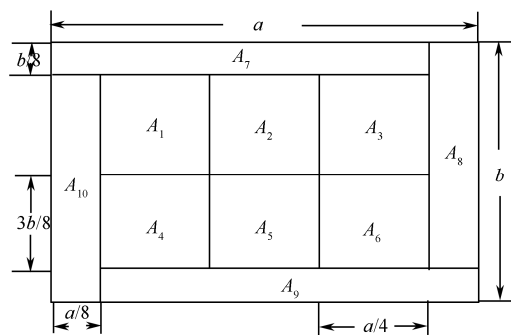


图 3.20 区域划分

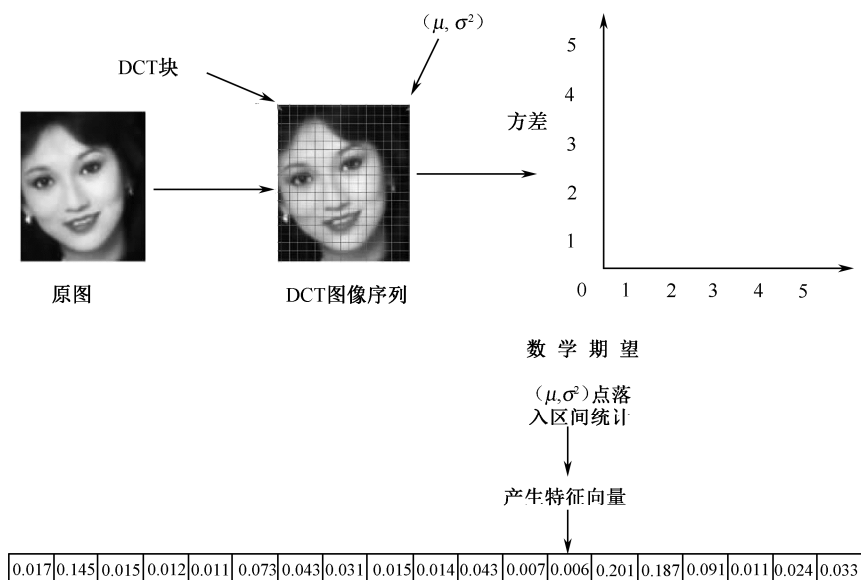


图 3.21 随机变量数字特征累加直方图提取过程

### 3. 边缘信息的提取

除纹理特征外, 边缘信息在表征图像内容上也具有重要意义。Abdelmalek<sup>[64]</sup>提出了利用 DCT 系数检测方向线的技术, 利用空域坡度为  $m$  的直线在 DCT 域中表现为坡度大约为  $1/m$  的直线这一特性, 在 DCT 域中提取水平、对角、垂直方向的特征信息进行图像的匹配。Shen<sup>[65]</sup>提出了一种从 JPEG 的 DCT 高频系数中检测出图像边缘和重要区域的方法, 其中, 边缘方向包括水平、垂直、对角、基本水平和基本垂直 5 种方向。实验结果表明, 基于 DCT 的边缘检测有着和 Sobel 边缘检测算子相媲美的性能。Lee<sup>[66]</sup>提出了一种利用边缘信息进行图像匹配的方法, 该方法首先从 AC 系数中提取二值边缘图, 然后利用边缘图计算边缘方向、边缘强度和边缘偏移量形成特征矢量, 最后按一定的相似性准则进行图像匹配, 实验结果表明该方法具有较好的边缘检测能力和检索效率。

黄祥林<sup>[67]</sup>等参考 3 级小波分解的情况 (如图 3.22 所示), 将 DCT 系数进行重组, 得到 DCT 域的多分辨率特征 (如图 3.23 所示)。

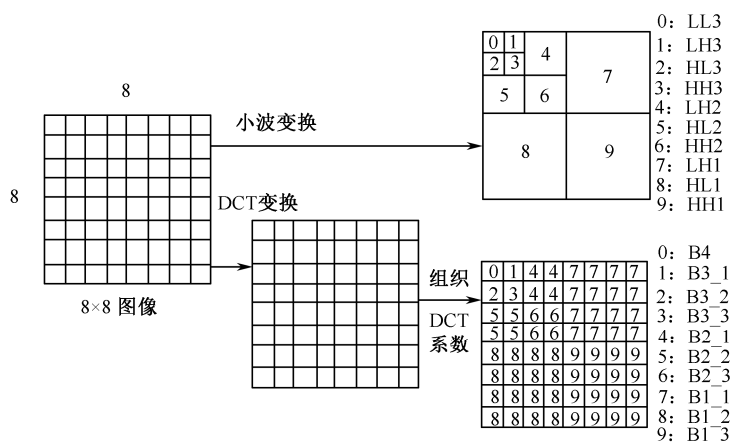


图 3.22 8×8 图像的小波变换与 DCT 变换

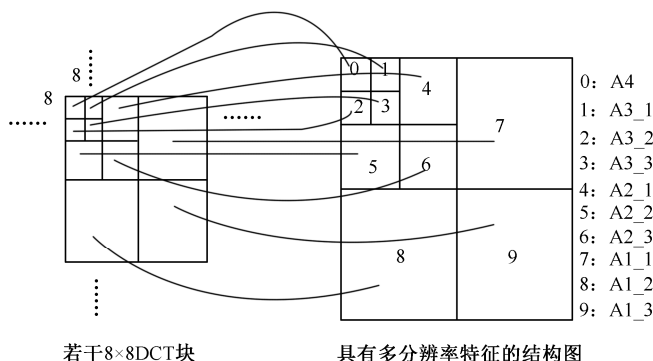


图 3.23 整幅图像中 DCT 块系数区域的组织

在此基础上, 设  $(x, y)$  表示图像区域中像素位置的坐标, 将  $A2\_1$ 、 $A2\_2$ 、 $A3\_3$  区域内同一位置的系数中的最大值组成一个与  $A2\_1$  区域同样大小的  $I(x, y)$  图像块, 即

$$I(x, y) = \max\{A2\_1(x, y), A2\_2(x, y), A2\_3(x, y)\} \quad (3-16)$$

然后对  $I(x, y)$  图像块进行简单的量化处理,  $M_1 = \max[I(x, y)]$ ,  $M_2 = \min[I(x, y)]$ ,  $M_3 = \text{Ave}[I(x, y)]$ ,  $M = (M_1 - M_2)/3$ ; 如果  $I(x, y) > M_3 + M$  或者  $I(x, y) < M_3 - M$ , 取  $I(x, y) = 255$ , 否则取  $I(x, y) = 0$ 。

经过这样对  $A2$  区域的处理就得到了图像的大致轮廓  $I(x, y)$ , 对  $A3$  区域也进行同样的轮廓提取, 而对  $A4$  区域 (即 DC 图) 则采用简单的 Roberts 算子提取轮廓。当得到图像的大致轮廓后, 进行基于轮廓的连通直方图的构造。对于给定分辨率为  $X \times Y$  的轮廓图像  $I(x, y)$ , 先计算其连通区域, 再计算具有相同连通区域面积 (用像素点表示) 的概率。连通直方图可通过下式计算。

$$H(k) = \sum_{I(x, y)} \delta^k \quad (3-17)$$

其中,  $\delta^k = \begin{cases} 1, & C = k \\ 0, & \text{其他} \end{cases}$ ;  $k$  表示连通区域的量化面积,  $k = 1, 2, \dots, M$ ,  $M \leq (X \times Y)/S$ ,  $S$

是连通区域面积的量化步长,  $C$  表示某个连通区域的量化面积。在实际应用中, 为消除图像尺寸的影响, 采用归一化的连通直方图, 即

$$H'(k) = H(k) / \sum_k H(k) \quad (3-18)$$

#### 4. 其他方法

除了上述的纹理特征和边缘信息的提取, 研究者还提出了其他的基于 DCT 域的图像检索方法。文献[68]定义某 AC 系数与该块中 DC 系数的比值为该系数的 A/D 能量, 在此基础上给出了一种基于 A/D 能量直方图的 JPEG 图像检索算法, 直接在 DCT 域中计算能量直方图。文献[69]分析了 DCT 块中系数分布的特点, 提出了一种 DCT 域中 MPEG-7 主色描述符的提取算法, 提高了压缩域图像进行特征提取的速度和效果。文献[6]直接在压缩域中利用 DCT 系数进行块分类, 每一类分块形成一个二值索引图, 统计该索引图的归一化转动惯量 (Normalized Moment Inertia, NMI) 值作为该类的一个特征, 所有类的 NMI 特征构成了图像的一个特征序列, 以此进行图像检索, 该方法不需要完全解压缩, 降低了计算复杂度, 对图像的平移、旋转和尺度变换有较好的鲁棒性。Shneier 等人<sup>[70]</sup>提出了一种基于 JPEG 压缩域的图像检索算法, 其主要思想是通过判定检索图像和目标图像中不相连区域对中 DCT 系数相似关系的大小进行检索, 先在图像内选取  $2K$  个互不相连的区域窗, 随机配对得到  $K$  个窗对, 对每个窗所包含的  $8 \times 8$  子块中每个系数进行均值计算, 得到 64 维的特征矢量, 在配对窗的特征矢量之间按其对应关系判定每一对分量的相似度大小并赋予该窗对一个比特 (0 或 1), 而检索图像和目标图像的相似与否取决于所有这些窗对比特流的相似度。Yu<sup>[71]</sup>提出一种

可用于直接比较两幅 JPEG 压缩图像相似度的测度,可直接利用 DCT 系数计算图像之间的相似程度。

Chang<sup>[72]</sup>等在 DCT 域中设计了一种提高检索效率的 JPEG 图像检索方法,称为 DCT 系数相关法。它根据 DC 系数和 AC 系数之间的相邻关系构造图像的索引特征,分别称为 DC 特征和 AC 特征,以此作为相似性检索的依据。具体特征的构造过程如图 3.24 所示。DC 特征是根据相邻块的 DC 系数的差值决定的一维 0、1 向量,差值大于等于 0,对应块被标记为 1,否则为 0,向量的长度等于块的总数。AC 特征是根据块中选取的 9 个 AC 系数相邻间的差值产生的,这些差值产生一个十进制索引值,由所有块的索引值产生一个直方图,它表示了一个索引值所对应的块数。两幅图像 DC 特征的比较采用异或运算,直方图的比较采用  $L_1$  距离方法,两种特征的相似性采用加权法确定最终的相似程度。这种方法在特征的构造上比较直接,运算量少,所以检索效率较高,但对于图像的旋转、缩放变换鲁棒性稍差。

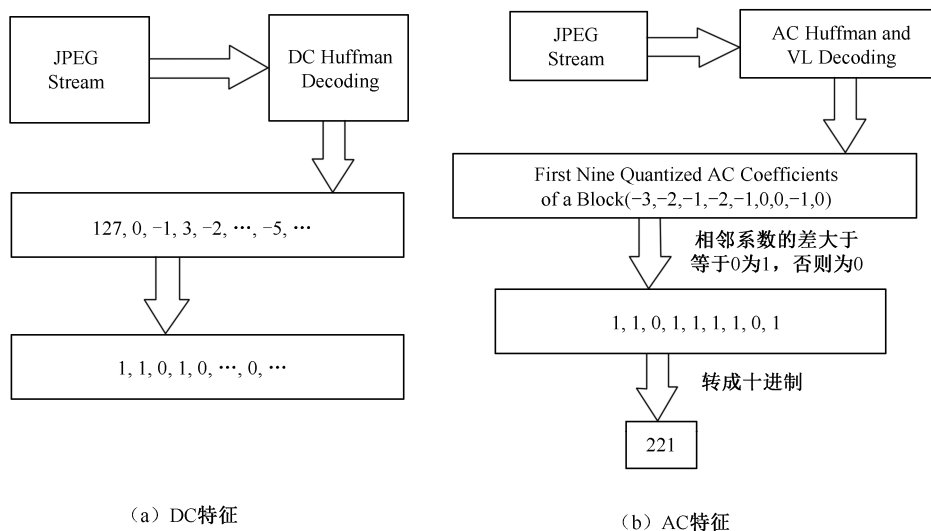


图 3.24 DCT 系数相关法的特征构造示意图

作为语义特征,图像中的字符对于理解图像/视频的内容有很大的帮助,要获得这种语义特征,需要先定位图像/视频中的字符区域。Zhong 等<sup>[73]</sup>提出一种 DCT 域字符区域定位方法,认为一行字符中字符与字符的间距对应于水平方向的频率变化,而多行字符之间的行距表现为垂直方向的频率变化,从而设计出一种两级字符区域定位方法:首先计算每一 DCT 子块的水平亮度变化信息,当亮度变化信息大于一定阈值时就认为该块为候选文本块,然后运用形态学法滤除孤立噪音块并合并并不相连的文本块,最后再用垂直亮度变化信息确认文本区域。黄祥林<sup>[74]</sup>等人也提出了一种 DCT 域的字符定位方法。这种方法依据字符具有的特殊线条结构而表现出明显的竖向、斜向、横向纹理特征的特点,计算每一 DCT 子块相应方向的频率变化程度,将其作为字符



的纹理特征并结合阈值法检测字符区。上述两种方法对尺寸较大且笔画为实心的字符区域及字符对比度较差的情况容易造成漏检，第一种情况是因为在 $8 \times 8$ 的子块中显示不出其频率的变化特性，后一种情况则可通过设计自适应阈值来解决。

### 3.3.3 基于小波压缩域

小波是一种处理多尺度可视化信息的强有力的数学工具，图像的小波表述则给出了图像在不同尺度下的变化信息<sup>[75]</sup>。小波母函数 $\psi(x)$ 是一个振荡衰减的函数（并且它在 $x$ 为无限远处收敛到零），对于一维信号函数 $f(x)$ ，设 $\psi_{2^j}(x) = 2^j \psi(2^j x)$ 是 $\psi(x)$ 在尺度 $2^j$ 的伸缩小波系数，设 $N$ 是 $f(x)$ 的采样点数，则 $f$ 在尺度 $2^j$ 的细节信号为 $f$ 与平移伸缩小波的内积，即

$$W_{2^j} f(n) = \langle f(u), \psi_{2^j}(u - 2^{-j}n) \rangle, \quad 0 \leq n \leq 2^j N \quad (3-19)$$

可以在不同的尺度 $2^j$ （即在尺度 $1/2, 1/4, \dots, 2^j, j \in \mathbf{Z}$ 且 $j \leq -1$ ）分别利用式（3-20）计算其小波系数，即

$$W_f = (W_{2^j} f)_{-J_{\max} \leq j \leq -1}, \quad J_{\max} = \log_2 N \quad (3-20)$$

接着用一种金字塔算法根据在尺度 $2^{j+1}$ 的小波系数来计算在尺度 $2^j$ 的小波系数，即

$$A_{2^j} f(n) = \sum_{k=-\infty}^{\infty} h(k - 2n) A_{2^{j+1}} f(k), \quad 0 \leq n \leq 2^j N \quad (3-21)$$

$$W_{2^j} f(n) = \sum_{k=-\infty}^{\infty} g(k - 2n) A_{2^{j+1}} f(k), \quad 0 \leq n \leq 2^j N \quad (3-22)$$

其中， $h$ 是尺度离散滤波器； $g$ 小波离散滤波器。

由于小波变换具有良好的时频局部性及与人眼视觉特性相符的多分辨率分析能力，因而它一经出现就被广泛地应用于图像压缩领域。近几年的研究表明基于小波变换的静止图像压缩系统在性能上优于基于DCT的压缩系统，从而促使小波变换在JPEG2000和MPEG24中获得应用。其主要优点可归纳为：①多分辨率性能；②较好的方向选择性；③高去相关性，以及高能量集中性；④没有块效应；⑤较好的人眼视觉特性。

目前，针对小波压缩域的图像检索技术已进行了多方面的研究。一部分学者致力于利用小波系数来进行纹理特征的提取。Smith<sup>[51]</sup>提出了一种纹理识别的方法，该方法首先是利用子带能量来定义纹理特征序列，然后对图像进行3级离散小波变换，并计算所有高频子带小波系数的幅度值，所有幅度系数子带通过上采样恢复到与原始图像同样的尺寸，共得到9个纹理通道，所有通道中位置相同的点构成一个9维矢量，对每一矢量中的分量进行二值量化处理，处理后的矢量共有 $2^9 = 512$ 种可能值，从而可

构建 512 级统计直方图,并用该直方图作为纹理特征来支持图像的检索。虽然这种方法具有较好的检索性,但是计算复杂度较高。Mandal 等<sup>[76]</sup>在此基础上提出了快速小波直方图检索方法,在基本不影响检索性能的情况下,大大降低了计算复杂度。这一类方法被统称为小波直方图法,是小波压缩域的典型算法之一,不但对纹理图像具有较好的检索效果,而且对自然图像也很有效。Jacobs 等人<sup>[77]</sup>基于小波变换系数直接实现了一种快速的多分辨率图像检索算法,在该算法中所有图像的尺寸都调整为  $128 \times 128$  并进行小波分解,对每一幅图像选出幅值最大的  $M$  ( $40 \sim 60$ ) 个小波系数记录其颜色、符号及索引位置,将这些数据组织起来作为表征图像的特征支持检索。虽然该文献给出了较好的检索性能,但由于特征数据中包含系数的位置索引,所以该算法明显不具备对平移和旋转等几何变化的鲁棒性<sup>[20]</sup>。Wang 等人<sup>[78]</sup>提出了另一种基于小波分解系数比较的图像检索算法。所有图像也调整为  $128 \times 128$  的尺寸并进行 4 级小波分解,利用最低一级的 4 个  $8 \times 8$  子带(LL、LH、HL、HH)图像进行 3 级检索:第一级通过 LL 子带系数方差的比较筛选出 20% 的图像,在第二级中对上一级筛选出的图像进行 LL 子带小波系数的直接比较,最后一级通过 LL、LH、HL、HH 4 个子带图像的系数的比较得到最终的检索输出图像。虽然该文献给出了比 Jacobs 等的算法更好的检索性能,然而该算法同样也不具备对平移、旋转的鲁棒性。闫允一<sup>[79]</sup>从提取兴趣点的角度,结合小波变换,提出一种基于稳定兴趣点和纹理特征的图像检索算法,首先用优化的 Hessian 检测器检测图像中的稳定兴趣点,计算稳定兴趣点的环形领域的伪泽尼克矩,伪泽尼克矩的抗噪性比泽尼克矩更强,并具有旋转不变性和鲁棒性。

基于小波压缩域的另一种典型方法是子带能量法。这种方法的基本思路是对原始图像进行小波分解,得到多个时频子带,计算每个子带的能量形成多维特征矢量用于检索。Chang<sup>[80]</sup>提出的通过不规则树分解来进行纹理分析的方法是子带能量法的雏形,在这种方法中通过计算  $J$  个中分辨率子带系数的能量形成  $J$  维特征矢量用于纹理匹配,取得了较好的效果。Lee 等在文献[81]中为了降低小波包分解时能量特征矢量的维数,仅仅选择包分解子带中能量较大的 7 个子带形成特征矢量用于纹理分类。实验表明,该方法在降低计算复杂度的同时,也可以取得很好的分类结果。Albanesi<sup>[82]</sup>提出了一种利用各子带之间的相关性进行图像检索的方法,首先进行小波分解,分别用各级的 LL 子带、LL+HL 子带、LL+LH 子带、LL+HH 子带近似重构上一分辨率图像,并计算它们与用 LL+HL+LH+HH 精确重构的图像之间的相关性,在每一层形成一个具有 4 个分量的特征矢量,用来进行图像分级检索。近年来,很多研究者通过选取不同的小波基(正交基、双正交基)、不同的小波分解方法(塔式小波分解、小波包分解等)及不同的子带能量计算法等对该类方法进行了深入研究。基于小波子带方面的研究还有文献[83]、[84]等。

与前述基于小波子带能量的方法不同,Mandal<sup>[85]</sup>提出了一种通过比较方向性子带的直方图来进行纹理匹配的方法。虽然不同图像在整个直方图上可能是相似的,但是不会具有类似的子带统计信息。借助不同级别的水平、垂直、对角信息的不同,可以

区别不同图像。在此基础上, Mandal 等人把小波分解的方向子带统计直方图特征用于图像检索<sup>[86]</sup>, 并构造了不同子带的矩特征用于支持小波域的图像检索。同时, 由于小波系数的分布近似符合广义高斯分布, Mandal 又提出对不同子带统计特性的匹配只需比较其用于描述分布的标准方差和形状参数, 从而实现了一种低复杂度、高效率的图像检索算法。Chen<sup>[87]</sup>则利用小波和隐含的马尔可夫模型提出了旋转和灰度不变的纹理图像分类技术, 首先从小波的各子带中提取灰度不变特征, 然后利用各子带序列组成的 Markov 模型捕获旋转的变化趋势, 这样能做到纹理匹配时的灰度、旋转不变。Wang 和 Mandal 等人又提出了以小波系数子带的统计特性为索引的技术, 每个子带小波系数的一阶矩(均值是其估计)和二阶矩(标准差是其估计)被当作图像索引。假设子带  $b$  包括  $k_b$  个系数,  $\mu_b$  和  $\sigma_b$  分别表示子带  $b$  相关系数的均值和标准差,  $\mu_b$  和  $\sigma_b$  为

$$\mu_b = \sum_{i=1}^{k_b} \frac{Wf_i}{k_b}, \quad \sigma_b^2 = \sum_{i=1}^{k_b} \frac{(Wf_i - \mu_b)^2}{k_b} \quad (3-23)$$

经过  $n$  级小波分解后, 小波系数分布在  $3n+1$  个子带中, 每个子带都有  $\mu_b$  和  $\sigma_b$  对, 总共  $3n+1$  个  $\mu_b$  和  $\sigma_b$  对被用于创建索引。

Sun 等<sup>[88]</sup>证明了任何一种对称小波基都可以作为边缘检测算子来实现图像的边缘检测, 只是采用反对称小波基时, 检测出的边缘图像会产生半个像素的平移。这为小波压缩域基于边缘轮廓的检索奠定了理论基础。双正交小波因其具有的良好特性(如线性相位、高阶消失矩等)被广泛应用于图像压缩领域。魏海等<sup>[89]</sup>依据反对称双正交小波具有的多尺度边缘提取能力, 在小波变换域内利用边缘像素的方向梯度相角信息构建直方图统计特征量来表征图像的内容, 提出了一种小波变换域内基于方向梯度相角直方图的图像检索算法。魏海等<sup>[90]</sup>证明了基于反对称双正交小波的塔式分解数据能够实现多尺度图像边缘提取。采用一定的反对称双正交小波对图像进行塔式分解, 并通过半重构和局部模极大值检测过程可以得到多个分辨率级上的边缘图像, 每个分辨率级上的边缘图像中的每个边缘像素都具有一个梯度模值和一个方向梯度相角值。假设第  $j$  级分辨率上的每一边缘像素点所具有模值为

$$M_j(x, y) = \left[ \left( \frac{\partial f_j(x, y)}{\partial x} \right)^2 + \left( \frac{\partial f_j(x, y)}{\partial y} \right)^2 \right]^{1/2} \quad (3-24)$$

方向梯度相角为

$$A_j(x, y) = \arctan \left[ \frac{\partial f_j(x, y)}{\partial y} / \frac{\partial f_j(x, y)}{\partial x} \right] \quad (3-25)$$

其中,  $f_j(x, y)$  为图像的第  $j$  级近似,  $x$  和  $y$  表示了边缘像素点在模图中的位置。利用每个分辨率级上的边缘图像进行方向梯度相角直方图的构造, 从而来表征图像的内容特性。理论分析和实验结果表明, 该算法不仅具有较高的检索效率, 同时具有较强的光照变化鲁棒性和一定程度的抗几何变化(尺度、平移、旋转等)的能力。

文献[91]也实现了一种基于小波和不变矩的形状检索方法,对原始图像进行小波变换,求取多尺度边界图像;计算每一边界图像的7个不变矩形成特征矢量并归一化;计算查询图像与目标图像特征矢量之间的距离来确定它们之间的相似性。实验表明,该方法能较好地描述图像的形状及空间分布信息,并具有较好的平移、尺度和旋转不变性。

文献[92]针对大尺寸图像特征提取算法复杂度高、特征信息容易缺失的问题,结合小波分解获得的小波系统,利用压缩感知理论中关于少量测量值可以精确重构原始信号的特性,提出了一种基于压缩感知的图像检索方法。首先对灰度图像进行二维小波分解,提取低频分量、水平分量、垂直分量和对角分量,并根据各分量图像的大小对分量图像进行缩放处理;然后对各分量图像进行分块预处理,构造分块多项式确定性测量矩阵,并把图像按从左到右、从上到下进行编号,计算测量系数,并对分块图像进行压缩感知快速测量,得到少量的压缩测量值代表图像的特征;最后采用加权距离方法计算图像测量值特征的相似度来实现图像检索。该方法为了克服随机测量矩阵的不足,采用分块多项式确定性测量矩阵对图像进行分块测量,加快了测量矩阵的构造,缩短了特征提取和匹配的时间,提高了图像的检索效率。

除上述这些直接基于小波变换系数提取各种表征图像内容特征的方法外,最近不少研究者开始研究是否能直接从压缩码流中提取表征图像内容的特征。因为小波零树编码方法被公认为是当前最好的基于小波变换的图像编码方法,JPEG2000中的熵编码部分就采用了零树编码的思想,所以针对小波零树编码算法研究者提出一些简单有效的检索技术。Wilson等<sup>[93]</sup>通过研究小波零树编码算法,统计EZW每一编码层的重要系数数目,并结合各层的阈值来计算各子带能量。这种方法不需要重建和存储整个小波分解矩阵,而是边扫描压缩符号流边计算特征矢量。理论分析和实验结果表明,这种方法可以在大大减少计算量和资源需求量的基础上得到较准确的子带能量特征。牛蕾等人<sup>[94]</sup>根据JPEG2000对感兴趣区域优先编码及感兴趣区域的形状可随意选取的特点,提出了一种在JPEG2000压缩码流不完全解码的情况下,实现多谱段遥感图像感兴趣目标的检索方法,该方法利用了遥感图像的性质,根据例子图像的谱特征对感兴趣区域的内容进行分析,并设计了一套相似性度量的方法。实验结果表明,此方法有较理想的图像检索效果和很高的检索效率,解决了应用上对实时性的高要求与遥感图像库数据海量性之间的矛盾。随着JPEG2000压缩标准的推出和普及,这种基于压缩符号流的检索技术将会得到更广泛的研究。

其他基于小波压缩域方面的研究还有:利用小波的极坐标表达方法研究其具有的旋转不变性<sup>[95]</sup>;研究非线性小波所具有的抗尺度、旋转、平移等几何不变的特性<sup>[96]</sup>;利用小波系数的局部极大值(模极大值)提取多尺度的边缘信息<sup>[97]</sup>;利用小波变换所具有的时频局域性进行特征点检测,然后基于这些感兴趣的特征点提取纹理、颜色特征实现图像检索<sup>[98,99]</sup>。

在JPEG2000标准没有发布之前,许多学者在小波压缩域中对图像检索技术进行

了大量的探讨。然而,在 JPEG2000 标准发布以后,在这个标准框架内探讨图像的检索技术更具有现实意义,许多学者在这方面已进行了不少的工作。Liu 等<sup>[100]</sup>在 JPEG2000 框架内给出了一种渐近位平面索引技术,他用小波系数的位平面重要位映射图及重要位直方图作为图像的索引,然后通过索引的匹配进行图像检索。他们还考虑了在压缩码流的包头信息中提取图像的特征<sup>[101]</sup>,虽然速度加快,但检索效果不好。Bhalod 等人<sup>[102]</sup>在该标准框架内结合感兴趣区域编码的特点给出了基于区域的图像检索方法,它通过 ROI 区域的跟踪和分析,提取 ROI 区域的轮廓及区域内的纹理和颜色特征作为图像检索的依据。Xiong 等<sup>[103]</sup>主要从节省内存的角度提出了一种基于块的图像索引方法,他提取的特征是块内的方差,但提供的信息量较少,效果不明显。Lin<sup>[104]</sup>给出了一种比较新颖的基于树距离比较的 JPEG2000 压缩域图像检索方法,它通过一种叫信息树的结构去描述小波系数的分布,然后通过信息树的距离比较进行图像检索。在分析 JPEG2000 压缩标准的基础上,直接利用 Taubman 优化截断嵌套块编码 EBCOT (Embedded Block Coding with Optimized Truncation) 后子图像块的“重要”或“不重要”信息,对图像进行分级匹配检索,用 0 和 1 分别表示 EBCOT 子像块的“不重要”和“重要”信息,则这些信息在小波变换的塔式分解方式中构成若干信息树,信息树中的每一个节点则对应一个 EBCOT 子像块,如图 3.25 所示。通过比较两幅图像的信息树,就可以在一定程度上得到其相似度,定义两幅图像的相似度为

$$S(I_1, I_2) = \sum_i a_i \sum_j \left[ D_T(T_{ij}^1, \bar{T}_{ij}^1) - D_T(T_{ij}^1, T_{ij}^2) \right] \quad (3-26)$$

其中,  $i$  表示对原始图像分块中第  $i$  个像块;  $a_i$  是对该像块的加权,当图像块中包含 ROI 时,可适当加大相应像块的权值;  $T_{ij}^1$  表示图像 1 的第  $i$  个像块的第  $j$  个信息树;  $\bar{T}_{ij}^1$  为对  $T_{ij}^1$  各节点取“非”;  $D_T(T_{ij}^1, T_{ij}^2)$  为两棵信息树的距离度量,定义为

$$D_T(T^1, T^2) = \sum_{s,k,l} \omega(s,k,l) \left[ T^1(s,k,l) \oplus T^2(s,k,l) \right] \quad (3-27)$$

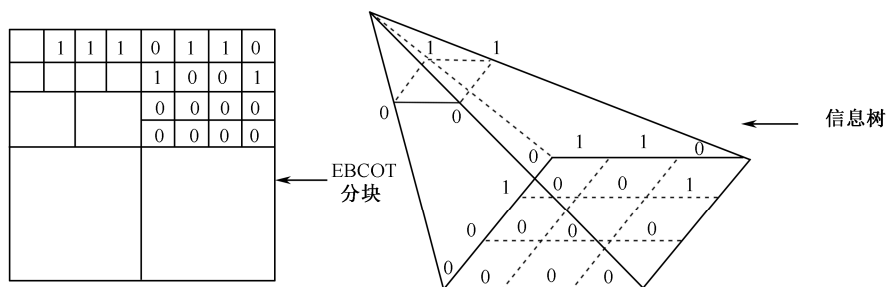


图 3.25 EBCOT 子像块的重要性判断和相应的信息树

其中,  $\omega(s,k,l)$  表示对应信息树中  $s$  尺度下的节点  $(k, l)$  的权值;  $T^1(s,k,l)$  表示信息树中  $s$  尺度下的节点  $(k, l)$  的信息比特。式 (3-27) 表示两棵信息树中的 0 和 1 分布越相似(对

应节点相同的 0 和 1 越多), 则两棵信息树的距离越小。

以上这些相关方法都是在小波压缩域内结合 JPEG2000 压缩特点进行图像检索的, 在检索速度上有了很大的提高, 但检索效果与传统的方法相比在整体上并没有太大的改进, 因此还需要进一步对该问题进行研究和探讨。

### 3.3.4 基于 K-L 变换域

K-L 变换是基于图像的统计特性的, 其基函数是图像自相关矩阵的特征矢量, 它具有最大的能量集中性, 是统计最优变换。设给定一组  $N$  个随机矢量

$$\mathbf{x} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_N] \quad (3-28)$$

其中, 每个矢量  $\mathbf{x}_i$  含  $M$  个分量, 即

$$\mathbf{x}_i = [x_{1i} \quad x_{2i} \quad \cdots \quad x_{Mi}]^T, \quad i=1, 2, \dots, N \quad (3-29)$$

这组随机矢量的均值矢量为

$$\mathbf{m}_x = E(\mathbf{x}) \quad (3-30)$$

其中,  $E[\cdot]$  代表期望值; 下标  $\mathbf{x}$  表示  $\mathbf{m}$  所对应的一组随机矢量。这组随机矢量的协方差矩阵为

$$\mathbf{C}_x = E[(\mathbf{x} - \mathbf{m}_x)(\mathbf{x} - \mathbf{m}_x)^T] \quad (3-31)$$

根据矩阵理论, 如果矩阵  $\mathbf{C}_x$  是一个实对称矩阵, 则总可以找到它的一组  $N$  个正交特征值。现令  $\mathbf{e}_i$  和  $\lambda_i$  ( $i=1, 2, \dots, N$ ) 分别为  $\mathbf{C}_x$  的特征矢量和对应的特征值, 并且这些特征值单调排列, 即  $\lambda_i \geq \lambda_{i+1}$  ( $i=1, 2, \dots, N-1$ )。再令  $\mathbf{A}$  为由  $\mathbf{C}_x$  的特征矢量组成其各行的矩阵, 并且  $\mathbf{A}$  的第一行为对应最大特征值的特征矢量,  $\mathbf{A}$  的最后一行为对应最小特征值的特征矢量。如果设  $\mathbf{A}$  是将  $\mathbf{x}$  转换为  $\mathbf{y}$  的变换矩阵, 则

$$\mathbf{y} = \mathbf{A}(\mathbf{x} - \mathbf{m}_x) \quad (3-32)$$

上式就称为 K-L 变换, 又称为霍特林变换, 由这个变换得到的  $\mathbf{y}$  矢量的均值为 0, 即

$$\mathbf{m}_y = 0 \quad (3-33)$$

$\mathbf{y}$  矢量的协方差矩阵可由  $\mathbf{A}$  和  $\mathbf{C}_x$  得到, 为

$$\mathbf{C}_y = \mathbf{A} \mathbf{C}_x \mathbf{A}^T \quad (3-34)$$

$\mathbf{C}_y$  是一个对角矩阵, 它的主对角线上的元素正是  $\mathbf{C}_x$  的特征值, 即

$$\mathbf{C}_y = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{bmatrix} \quad (3-35)$$

它的主对角线以外的元素均为零, 这表明  $\mathbf{y}$  矢量的各元素是不相关的。考虑到  $\lambda_i$  也是

$C_x$  的特征值, 并且沿  $C_x$  对角矩阵的主对角线上的元素是  $C_x$  的特征值, 所以可知  $C_x$  和  $C_y$  具有相同的特征值和特征矢量<sup>[105]</sup>。

从 K-L 变换的定义可以看出, 由于其基函数是自适应的, 因此可以通过映射图像到 K-L 空间, 然后比较 K-L 变换系数来实现图像检索。该变换具有很好的图像检索能力, 在人脸的识别上得到了较好的效果。Pentland 等利用 K-L 变换给出了一种人脸识别技术<sup>[106]</sup>。首先, 随机抽取一组人脸图像, 根据这组图像计算最优变换矩阵; 然后, 利用该矩阵计算查询图像和目标图像的 K-L 变换系数; 最后, 计算二者 K-L 变换系数之间的欧式距离, 并以此距离进行人脸识别。因为 K-L 变换系数只有最前面的系数比较大, 所以可以利用前面的少数系数, 这样就可以降低特征矢量的维数, 减少图像匹配的计算量。

虽然 K-L 变换提取的是一幅图像的最主要特征, 但常常携带一些与目标识别无关的信息, 如光照方向、光照强度等, 即使增大样本特征的数量也无法消除这些无关信息。针对这一问题, Swets<sup>[107]</sup>提出了一种 DKLT (Discriminant K-L Transform), 主要思路是对图像先进行 K-L 变换, 然后对变换的结果进行区别分析, 从而得到一组最具有区别能力的特征。在 DKLT 中, 不同类之间的距离被最大化, 而同类之间的距离被最小化。实验结果表明, 该方法在检索效率上要优于 K-L 变换方法, 但计算量几乎是 K-L 变换的 2 倍。另外, 该方法还需要人工标注每幅训练图像类别, 只能适用于实时性不高的场合, 如人脸识别等, 而对于网上查询、大容量数据库检索等是不适合的。曹奎等人<sup>[108]</sup>通过对彩色空间的分析, 提取图像中的颜色不变量, 然后在频域内对这样的颜色信息进行分析, 并将频域分析的结果进行 K-L 变换, 变换后的低维向量即为图像的颜色表示, 从而给出了一种描述图像视觉特征的图像表示方法, 并据此计算图像之间的全局相似度, 在此基础上, 讨论了图像的相似度量及相应的图像检索技术, 并给出了实验结果和图像检索性能的评价。杨琼等人<sup>[109]</sup>提出对称主分量分析, 该算法首先引入镜像变换, 生成镜像样本, 然后依据奇偶分解原理, 生成镜像奇、偶对称样本, 并分别进行 K-L 展开, 提取镜像奇、偶对称 K-L 特征分量, 最后根据奇、偶对称 K-L 特征分量在人脸中所占能量比例的不同及对视角、旋转、光照等干扰的不同敏感程度进行特征选择, 节省计算与存储开销, 增强算法的实用性能。文献[110]提出了一种新的遥感图像检索方法, 首先, 对图像进行主成分变换, 对变换后的第一主成分图像进行五叉树分解, 将大幅面的遥感图像分成一系列的子图像; 然后, 利用多通道 Gabor 滤波器与子图像进行卷积运算, 提取其纹理特征, 同时计算像元值的方差和三阶矩作为各子图像的色调特征; 最后, 以子图像为特征基元, 构建图像的色调直方图和纹理直方图, 以多特征直方图匹配方法计算图像相似度实现遥感图像检索。到目前为止, 由于其固有的计算复杂性, K-L 变换很少被用于传统的图像压缩; 然而, K-L 变换在多波段图像的分析处理和编码中应用却比较广泛。

### 3.4 空间域和变换域的融合检索

基于空间域和基于变换域的图像检索方法各有优劣。一种检索方法是否能得到普及应用,在很大程度上取决于其所基于的压缩方法的性能和使用范围。K-L 变换虽然统计最优,但其固有的计算复杂性阻碍了本身及与之相关的检索技术的广泛应用。DFT 在图像分析中具有重要的作用,但 DFT 对非周期图像信号压缩能力不是很理想,很少被用于图像压缩,其检索能力也就无从谈起。矢量量化的编码过程本身就是一种索引机制,最易实现压缩域图像检索,但矢量量化的编/解码过程很不对称。分形虽在编码方面具有很大的潜能,但是分形码是高度非线性且依赖于待编码图像的,只有图像本身具有明显的自相似性或统计相似性时,才能获得很高的压缩率,另外分形的编/解码过程也很不对称。由于这些原因,基于 DFT、K-L 变换、矢量量化和分形等的图像检索技术在人们进行了一些探索性研究后,并未成为压缩域检索技术的研究主流,不过在一些特殊的领域,如多波段遥感图像、医学图像,由于要根据图像自身的特点选取最合适的压缩方法,故这些检索技术仍然有其应用潜力。为了充分利用各种技术的优点,一些专家们致力于研究空间域和变换域相融合的图像检索技术,取得了较好的成果。

Idris<sup>[111]</sup>提出了一种在小波域利用矢量量化的图像检索技术,根据小波各子带的特性,对其系数进行矢量量化,并利用其结果进行图像比较。而 Swanson 等<sup>[112]</sup>设计的一种基于分割的图像编码算法的主要思想是,对于所分割的物体对象利用小波、矢量量化的方法编码以提供可直接面向对象的图像索引能力,而对残差图像则进行基本的 JPEG 编码。Podilchuk<sup>[113]</sup>在 DCT 压缩域采用矢量量化的方法对 DCT 系数进行量化,并利用矢量量化的索引技术进行人脸识别。

Swanson<sup>[114]</sup>等提出一种基于小波域的 VQ 检索方法,它不同于传统的基于内容检索将特征矢量单独存储为特征矢量数据库的方法,而是将特征矢量存储为压缩图像的一部分,大大降低了检索复杂度对存储空间的要求。算法首先采用一定分割方法在图像中找到感兴趣的物体,而后对物体进行小波变换和嵌入式矢量量化,其他的非物体图像区域使用简单的编码算法,如 JPEG、小波或矢量量化编码,如图 3.26 所示。其中,多分辨率头文件采用从低分辨率到高分辨率的组织结构,检索过程利用检索图像的有限 VQ 字典得到的适当码字进行,开始寻找各幅图像最低分辨率下的码字,如果匹配,则说明在该图像中找到在此尺度下和检索图像相似的物体,而后可以在更高分辨率下对码字进行匹配,逐步找到相似图像。



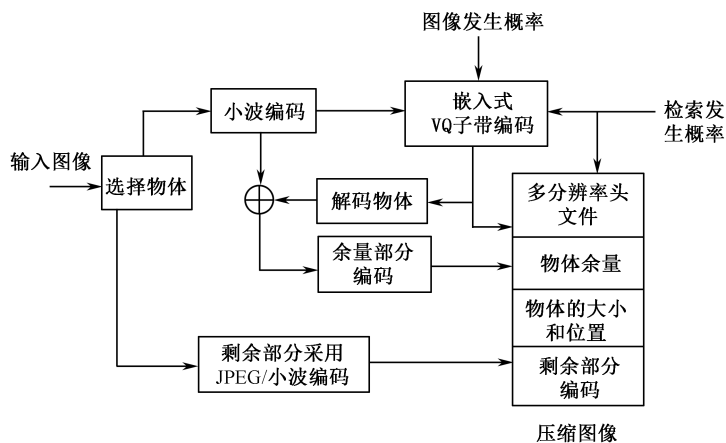


图 3.26 小波域的 VQ 检索方法示意图

### 3.5 DCT 压缩域内的纹理特征

传统的基于 JPEG 图像的检索首先要对压缩图像解压缩，而后利用基于像素域的图像检索方法提取颜色、纹理、形状和空间位置关系等特征，采用一定的相似性度量准则进行检索。这些方法取得了很好的效果，但也存在一些缺点，如它们使用的特征种类比较单一，一般只单独利用纹理或者颜色特征，没有充分利用 DCT 系数所包含的丰富信息；特征的组织形式比较简单，大都采用直方图或随机块的形式，没有充分利用图像的结构性信息。而 DCT 系数有其内在的优良性质，多数统计特征可直接通过图像子块 DCT 直接统计，这些统计特征又直接反映了图像的视觉特征，比在空间域进行类似特征统计分析大大缩短了运行时间。

从图 3.3 和图 3.11 可以看出，基于 JPEG 压缩域的图像检索数据源可以来自 JPEG 解压过程的任意环节，针对不同的数据源格式和具体应用，可以有不同的检索模式。由于熵编码属于非结构化、非字节对齐编码，在熵编码后进行特征提取操作十分困难，所以现有的检索方法通常都在熵解码后或熵解码和反量化后进行特征提取。而且，如果适当组织 DCT 系数，使之反映出图像纹理的方向性，则有利于利用 DCT 系数进行图像检索。本节首先对 JPEG 图像进行熵解码得到量化的 DCT 系数，然后在 DCT 压缩域中，利用 DCT 系数的分布特点，介绍一种图像纹理特征的提取算法——复杂度直方图方法<sup>[115]</sup>。该方法不受旋转变换、平移变化等的影响。同时，考虑到 DCT 系数在块中的空间位置分布的不同对最后检索效果的影响，选取每个 DCT 块中能量最大的 9 个系数作为重要系数，以此来对每个块的复杂度设置权值，从而避免了由于复杂度相同而系数空间分布信息不同而造成的误检和漏检情况。该方法方便简单，直接在

DCT 域提取图像的特征矢量, 计算复杂度低, 而且能很好地反映图像中的纹理分布。同时, 还可以根据 DCT 块各系数所处位置的空间信息改进复杂度直方图, 进一步提高图像的检索效率。

### 3.5.1 复杂度的定义

近年来, 在生物医学信号处理, 特别是脑电分析中, 复杂度分析引起了人们广泛的关注<sup>[116]</sup>。复杂度方法是非线性动力系统的一种度量方法, 它是作用于时间序列的一种指标。一段时间序列通过一个给定的复杂度算法处理, 可以得到一个复杂度值, 这个值表示的是序列的非规则程度, 不规则变化序列的复杂度值较高, 常数序列、周期序列这种模式比较简单的序列的复杂度值就比较低, 这里指的是一维复杂度。文献[117]提出了一种 CO 复杂度的概念, 主要思想是把信号分解成规则成分和非规则成分两个部分, CO 复杂度定义为非规则部分在原信号里所占的比例。复旦大学数学系的蔡志杰严格证明了 CO 复杂度的一些性质<sup>[118]</sup>, 表明 CO 复杂度在一定条件下可以作为时间序列随机程度的指标, 在随机性复杂度的意义下也可作为复杂度的一个定量指标。但由于二维空间与一维空间存在着根本的区别, 一维的符号序列是自然的和明确的, 可以从左或者从右给序列中的每个符号编上序号, 而二维的符号矩阵没有自然的顺序, 将每个符号排序的方法也不是唯一的, 因此采用一维复杂度的定义无法很好地描述二维空间的数据分布信息。文献[118]提出了一种用于度量图像等二维结构的复杂度算法, 定义如下。

设  $\{f(j, k), j = 0, 1, 2, \dots, M-1, k = 0, 1, 2, \dots, N-1\}$  是一个长度为  $M$ 、宽度为  $N$  的平面图像序列,  $\{F_{MN}(m, n), m = 0, 1, 2, \dots, M-1, n = 0, 1, 2, \dots, N-1\}$  为相应的二维离散傅里叶变换序列, 定义为

$$F_{MN}(m, n) = \frac{1}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j, k) e^{-2\pi i \frac{jm}{M}} e^{-2\pi i \frac{kn}{N}}, m = 0, 1, 2, \dots, M-1, n = 0, 1, 2, \dots, N-1 \quad (3-36)$$

其中,  $i = \sqrt{-1}$  是虚数单位, 为了书写方便, 记  $W_M = e^{\frac{2\pi i}{M}}, W_N = e^{\frac{2\pi i}{N}}$ , 则  $F_{MN}(m, n)$  可以简写为

$$F_{MN}(m, n) = \frac{1}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j, k) W_M^{-jm} W_N^{-kn}, m = 0, 1, 2, \dots, M-1, n = 0, 1, 2, \dots, N-1 \quad (3-37)$$

记  $\{F_{MN}(m, n), m = 0, 1, 2, \dots, M-1, n = 0, 1, 2, \dots, N-1\}$  的均方值为  $G_{MN}$ , 则

$$G_{MN} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |F_{MN}(m, n)|^2 \quad (3-38)$$

定义

$$\tilde{F}_{MN}(m,n) = \begin{cases} F_{MN}(m,n), & |F_{MN}(m,n)|^2 > G_{MN} \\ 0, & |F_{MN}(m,n)|^2 \leq G_{MN} \end{cases} \quad (3-39)$$

对  $\{\tilde{F}_{MN}(m,n), m=0,1,2,\dots,M-1, n=0,1,2,\dots,N-1\}$  作离散傅里叶逆变换, 得到一个新的二维平面图像序列

$$\tilde{f}(j,k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \tilde{F}_{MN}(m,n) W_M^{jm} W_N^{kn}, \quad j=0,1,2,\dots,M-1, k=0,1,2,\dots,N-1 \quad (3-40)$$

二维复杂度可定义为

$$C_0 = \frac{\sum_{j=0}^{M-1} \sum_{k=0}^{N-1} |f(j,k) - \tilde{f}(j,k)|^2}{\sum_{j=0}^{M-1} \sum_{k=0}^{N-1} |f(j,k)|^2} \quad (3-41)$$

蔡志杰证明了二维复杂度具有下面的性质<sup>[119]</sup>。

- (1) 如果图像是单一颜色的(常数二维结构), 它的二维复杂度为 0。
- (2) 如果图像在纵向或横向是周期的, 随着图像不断延伸, 二维复杂度趋向于 0。
- (3) 对于任意图像, 它的二维复杂度总介于 0 和 1 之间。如果图像是随机的, 每个像素的取值服从于独立同分布, 且有有限四阶矩, 记其均值为  $Ef = \mu$ , 方差为  $Df = \sigma^2$ , 则当  $M \rightarrow \infty$  或  $N \rightarrow \infty$  时, 二维复杂度以概率 1 收敛于  $\sigma^2 / (\mu^2 + \sigma^2)$ , 特别地, 当  $\mu = 0$  时, 二维复杂度以概率 1 收敛于 1。

从定义可以看出, 二维复杂度利用一个阈值将序列的傅里叶变换分为规则和不规则两部分, 最后计算非规则部分占原信号的比例, 从而通过频谱来确定图像的不规则程度。显然图像越不规则, 越复杂, 二维复杂度越大。文献[118]证明, 二维复杂度能很好地反映图像在直观上体现出的规则和非规则程度, 可以作为图像等二维结构非规则程度的度量。

### 3.5.2 复杂度直方图

JPEG 压缩标准中, DCT 的结果使得每一个  $8 \times 8$  的 DCT 块都有 1 个 DC 系数和 63 个 AC 系数。而对一个 DCT 块来说, 原始图像中纹理复杂的 DCT 块, 系数分布比较复杂, 中高频系数非零值较多, 而平滑的 DCT 块中, 系数分布较简单, 中高频系数非零值较少。因此, DCT 块的二维复杂度也可以在一定程度上间接反映原始图像块中的纹理信息。因此, 在算法中, 我们提取每个 DCT 块的复杂度构造复杂度直方图来描述原图像的纹理特征<sup>[120]</sup>。

假设原图像  $I$  的大小为  $M \times N$ , JPEG 压缩图像采用  $8 \times 8$  大小分块, 整幅图像共分子块数目为  $\frac{M}{8} \times \frac{N}{8}$ , 用  $C_{mn}$  来标记  $(m,n)$  处子块的复杂度, 则复杂度直方图可定义为

$$H_c(I) = \frac{64}{M \times N} \Pr[C_{mn} = c], \quad c \in [0, t] \quad (3-42)$$

其中,  $m = 0, 1, 2, \dots, \frac{M}{8} - 1$ ;  $n = 0, 1, 2, \dots, \frac{N}{8} - 1$ ;  $t$  为复杂度的量化级数。

用上述算法提取的直方图在一定程度上反映了原始图像中的纹理信息, 但复杂度直方图同颜色直方图一样, 仅仅体现了 DCT 系数的统计分布, 而没有包含其空间分布信息。而对每个块中的 DCT 系数来说, 不同位置的系数反映了原始图像中不同的方向信息, 如果图像块在水平、垂直、对角方向具有明显的边缘特征, 那么其变换后的 DCT 系数将相应地在水平、垂直和对角方向上的值较大<sup>[64]</sup>。因此, 如果不考虑 DCT 系数在块中的分布, 仅仅采用复杂度直方图用于检索, 可能会造成误检或漏检。

考虑到这个因素对最终检索结果的影响, 在算法中, 为体现 DCT 系数的空间分布信息, 利用块中能量最大的前几个 AC 系数来为每个块的复杂度引入权函数。利用能量最大的几个 AC 系数的分布来设定权函数, 是因为它们保持了大量的能量和纹理信息, 能从另一方面来体现块的特征。由于在 DCT 块中与 AC 系数对应的频率越高, 量化效应对它的影响就越大, 其带来的误差也越大, 所以使用少数 AC 系数比使用较多 AC 系数要好, 而且在整幅图像中有小物体平移时, 这种方法也不受影响, 同时也是出于计算量的考虑, 选取了前 9 个能量最大的系数。如图 3.27 所示, 虽然经过 DCT 变换后, 低频能量都集中在了图像的左上角, 但是前几个能量最大的低频系数并非如图 3.27 (b) 所示的排列方式, 当任取图 3.27 (a) 中的一个子块做实验时, 发现其能量最大的 9 个低频系数的位置如图 3.27 (c) 所示。同时, 对具有相同或相近复杂度的 DCT 块来说, 考虑到如果块中各分量的位置不同, 在对 DCT 系数进行排序后, 各系数的移动次数也将不同, 因此采用排序时各分量移动次数的差异可有效地对 DCT 块进行区分。为此采用类似式 (2-27) 的方法, 按如图 3.6 所示的 Z 字形扫描方式将前 32 个系数进行排序 (许多高频系数已经被量化成 0, 由于计算量的考虑, 故将它们忽略), 得到其中能量最大的 9 个系数, 引入加权函数  $f(x)$ , 即

$$f(x) = 1 + \frac{m_x}{m_{\max}} \quad (3-43)$$

其中,  $x$  表示任意的 DCT 块;  $m_x$  表示 DCT 块  $x$  经过排序后, 前 9 个能量最大的系数移动的总次数;  $m_{\max}$  表示排序时, 9 个系数需要移动的最大次数, 即对逆序 DCT 系数进行排序时 9 个系数需要移动的总次数。

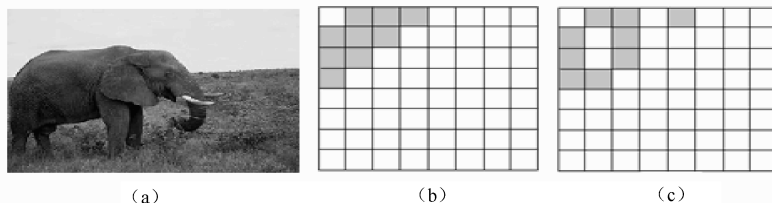


图 3.27 图像及其中某个 DCT 块中最大能量系数的分布图

为了选取合适的量化策略，我们对复杂度的分布进行了统计，首先从 Corel 图像库中随机抽取 200 幅图像组成测试库，所选取的图像包括多个类型，这样可以使统计结果具有一般性，并假定选取的图像库足够大，测试库中图像的分布特征足以能描述整个图像集的分布特征，然后对所有图像的 DCT 块计算其复杂度并进行统计，以此作为量化的依据。统计结果如图 3.28 所示，由图可知，DCT 块的复杂度分布不均匀，大多数复杂度值位于区间 $[0.03, 0.5]$ 。因此，我们在算法中采用如下非均匀量化的方法来构造复杂度直方图：在区间 $[0, 0.03]$ 上将复杂度值均匀量化为 10 柄，在区间 $[0.03, 0.5]$ 上量化为 120 柄，在区间 $[0.5, 1]$ 上量化为 40 柄，总计 170 柄<sup>[121]</sup>。文献[118]已经证明，图像的大小和形状改变时对最终复杂度的计算没有影响，而直方图统计特性通常对图像的旋转和平移等变换具有较好的鲁棒性，归一化后具有一定的尺度不变性，因此该算法对原始图像的各种形变具有一定的鲁棒性。算法中通过除图像尺寸大小来进行归一化。

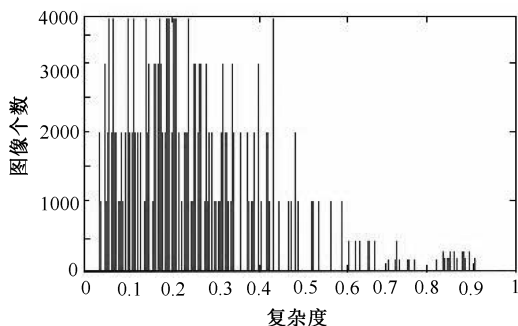


图 3.28 测试库图像 DCT 块分布统计规律

### 3.6 DCT 压缩域内的形状特征

上一节讨论的特征仅仅利用了 DCT 系数所隐含的纹理信息，因此对纹理丰富的图像检索效果较好。本节进一步探索 DCT 系数区域所具有的特性，可以利用 DCT 块的系数分布特征来提取图像的边缘及形状信息<sup>[122]</sup>。

#### 3.6.1 理想边缘模型 DCT 块的分类

从 DCT 的定义可以看出，变换后每一个  $8 \times 8$  的 DCT 块中的系数都是块内所有像素值的线性组合。以  $AC_{01}$  为例，有

$$AC_{01} = \frac{C_0 C_1}{4} \left\{ \cos \frac{\pi}{16} \left[ \sum_{j=0}^7 f(0, j) - \sum_{j=0}^7 f(7, j) \right] + \cos \frac{3\pi}{16} \left[ \sum_{j=0}^7 f(1, j) - \sum_{j=0}^7 f(6, j) \right] \right. \\ \left. + \cos \frac{5\pi}{16} \left[ \sum_{j=0}^7 f(2, j) - \sum_{j=0}^7 f(5, j) \right] + \cos \frac{7\pi}{16} \left[ \sum_{j=0}^7 f(3, j) - \sum_{j=0}^7 f(4, j) \right] \right\} \quad (3-44)$$

从式 (3-44) 可以看出,  $AC_{01}$  取决于 DCT 变换前  $8 \times 8$  块的左右两部分的亮度值的差, 体现了图像在垂直方向上的灰度差。同理,  $AC_{10}$ 、 $AC_{20}$ 、 $AC_{02}$ 、 $AC_{11}$  等几个系数也分别体现了图像在不同方向上分布的灰度差, 其物理意义如图 3.29 所示。

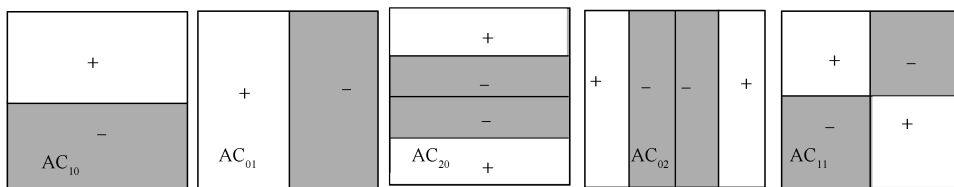


图 3.29 DCT 块中 5 个 AC 系数的物理意义

在上述分析的基础上, 文献[65]根据当只有一条边缘线穿越 DCT 块时可能出现的情况提出了两种理想边缘模型, 如图 3.30 所示,  $\theta$  表示边缘方向的角度值,  $d$  表示边缘相对中心的偏移,  $I$  是指边缘强度。对于图像经过 DCT 变换后得到的所有 DCT 块, 如果直接在 DCT 压缩域, 通过 AC 系数的计算完全表示出其边缘模型, 有两个问题。一方面, 要准确地表示出边缘的方向和相对偏移, 才能比较准确地表示出边缘, 这涉及较多的 AC 系数和很大的计算量; 另一方面, 要完全表示出边缘的二值模型, 需要用 64 个数据单位描述每个块的边缘, 这为以后计算增加了负担。针对以上问题, 文献[65]提出了一种简化的边缘模型。对于每个  $8 \times 8$  的块, 仅仅通过使用前 5 个 AC 系数数值的大小, 就可以判断出边缘的大概方向, 从而可以将其划分为如图 3.31 所示的 20 种边缘类型。

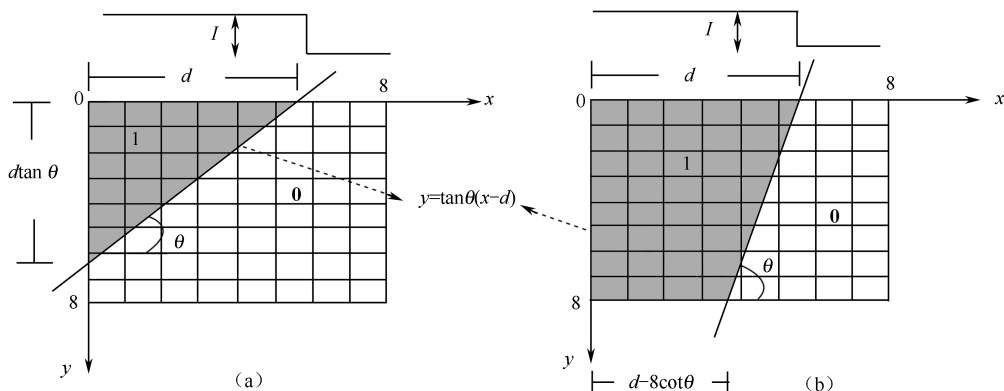


图 3.30 两种理想边缘模型

$AC_{10} > 0$ $AC_{01} > 0$	$AC_{20} = 0$		(b)	$AC_{10} > 0$ $AC_{01} < 0$	$AC_{20} = 0$		(b)
	$AC_{02} = 0$		(b)		$AC_{20} = 0$		(b)
	$AC_{11} > 0$		(a)		$AC_{11} > 0$		(a)
	$AC_{11} < 0$		(a)		$AC_{11} < 0$		(a)
$AC_{10} < 0$ $AC_{01} < 0$	$AC_{20} = 0$		(b)	$AC_{10} < 0$ $AC_{01} > 0$	$AC_{20} = 0$		(b)
	$AC_{02} = 0$		(b)		$AC_{02} = 0$		(b)
	$AC_{11} > 0$		(a)		$AC_{11} > 0$		(a)
	$AC_{11} < 0$		(a)		$AC_{11} < 0$		(a)
$AC_{10} = 0$	$AC_{01} > 0$		(b)	$AC_{10} \neq 0$	$AC_{01} > 0$		(b)
$AC_{10} = 0$	$AC_{01} < 0$		(b)	$AC_{10} \neq 0$	$AC_{01} < 0$		(b)

图 3.31 利用 DCT 块中 AC 系数之间的关系提取的边缘类型

### 3.6.2 空间边缘分布特征的提取

定义了 DCT 块的边缘类型后,一幅  $M \times N$  的图像就对应着一个  $(M/8) \times (N/8)$  的矩阵  $\mathbf{P}$ , 其中  $\mathbf{P}(x, y)$  的值为  $(x, y)$  处 DCT 块所属的边缘类型索引值。为了提取边缘分布的空间信息,我们针对  $\mathbf{P}(x, y)$  中的某一类索引值,保留该索引值的位置上的值,将其余的位置置 0, 构成一个该类的空间分布图,在此基础上提取边缘的空间分布特征<sup>[120]</sup>。

设  $\mathbf{A}_i = \{(x, y) | (x, y) \in \mathbf{P}, \mathbf{P}(x, y) = i, 1 \leq i \leq 20\}$  表示  $\mathbf{P}$  中索引值为  $i$  的所有点的集合。设  $|\mathbf{A}_i|$  表示集合  $\mathbf{A}_i$  中点的数目,  $C_i = (x_i, y_i)$  为  $\mathbf{P}$  中索引值为  $i$  的所有点的质心。 $x_i$  和  $y_i$  定义同式 (2-39)。

设  $r_i$  表示  $\mathbf{P}$  中  $(x, y)$  处索引值为  $i$  的点同其质心的距离, 其定义为

$$r_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (3-45)$$

则  $\mathbf{P}$  中所有属于  $i$  的点到质心的距离和为

$$R_i = \sum r_i = \sum_{(x, y) \in \mathbf{A}_i} \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (3-46)$$

算法中,将每一种边缘类型构成的空间分布图中所有点到其质心的距离和作为表征其空间分布的特征,从而构造了整个图像的空间分布特征  $(R_1, R_2, R_3, \dots, R_{20})$ 。

由于各个 DCT 块的边缘类型是利用每个 DCT 块的 5 个系数粗略分类的,因此,即使相同类型的 DCT 块,其能量也有可能不同。所以,对两幅图像而言,当两者对应边缘类型的空间分布一致时,由于每类边缘的能量的大小不一,也会造成图像的误检。考虑到这个因素对最后检索效果的影响,我们以每类边缘空间分布图上值不为 0

的点所代表的 DCT 块的边缘强度总和作为图像的另一特征。对于图 3.31 中 20 种边缘, 可以利用如图 3.32 所示的对称准则<sup>[66]</sup>将其转换为图 3.30 中 (a) 和 (b) 两种理想边缘, 然后计算其边缘强度和。强度的数学定义如下, 具体推导过程见文献<sup>[66]</sup>。

$$I_{(a)} = 2\sqrt{2}\left(\frac{\pi}{8}\right)^2 \frac{AC_{10}AC_{01}}{\sqrt{(AC_{10} - AC_{20})(AC_{01} - AC_{02})}} \quad (3-47)$$

$$I_{(b)} = 2\sqrt{2}\left(\frac{\pi}{8}\right)^2 AC_{01} \sqrt{\frac{\sqrt{2}AC_{10}AC_{11}}{\sqrt{2}AC_{10}AC_{11} - AC_{20}AC_{01}}} \quad (3-48)$$

则边缘强度和

$$E_i = \begin{cases} \sum I_{(a)}, & i \in (a) \\ \sum I_{(b)}, & i \in (b) \end{cases} \quad (3-49)$$

则提取的边缘强度特征可表示为  $(E_1, E_2, \dots, E_{20})$ 。

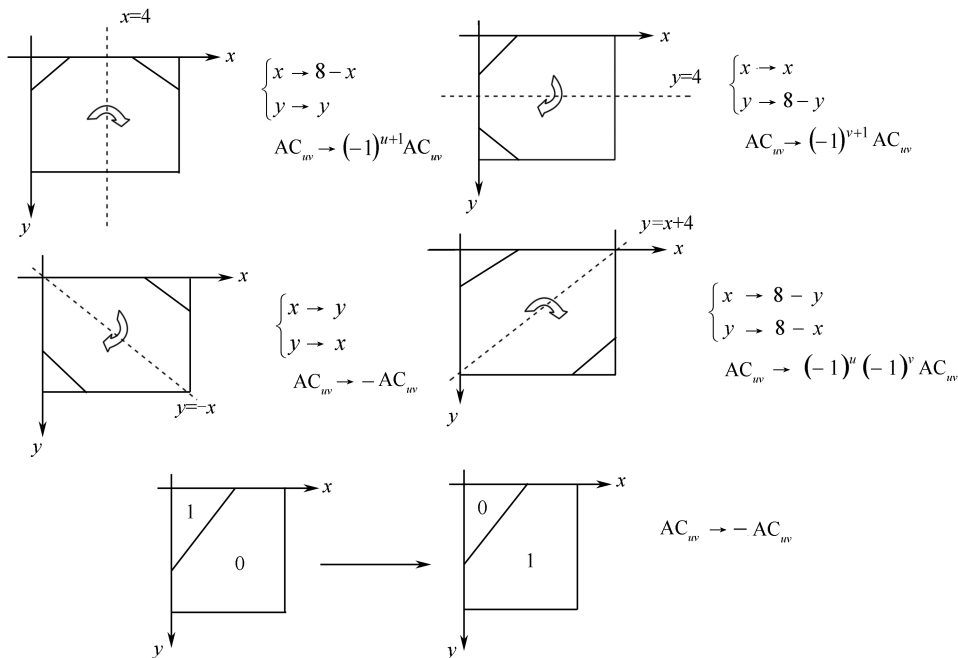


图 3.32 计算 AC 系数的对称准则

## 参 考 文 献

- [1] 邵虹. 基于内容的医学图像检索关键技术研究[D]. 沈阳: 东北大学, 2004.
- [2] 赵珊. 基于内容的图像检索关键技术研究[D]. 西安: 西安电子科技大学, 2007.



- [3] 沈兰荪, 张菁, 李晓光. 图像检索与压缩域处理技术的研究[M]. 北京: 人民邮电出版社, 2008.
- [4] Wu D, Tan E C. Comparison of lossless image compression algorithms[J]. IEEE TENCON, 1999,6:718-721.
- [5] 赵德斌, 陈耀强, 高文. 基于块方向预测和 Context 的图像无失真编码方法[J]. 软件学报, 1998, 9(10):765-769.
- [6] 张问银. 压缩域图像检索及汉字数学表达式研究[D]. 成都: 中国科学院成都计算机应用研究所, 2005.
- [7] 沈兰荪, 张菁, 李晓光. 图像检索与压缩域处理技术的研究[M]. 北京: 人民邮电出版社, 2008.
- [8] 沈兰荪. 图像编码与异步传输[M]. 北京: 人民邮电出版社, 1998.
- [9] Ahalt S C, Krishnamurthy A K, Chen P, et al. Competitive learning algorithms for vector quantization[J]. Neural Networks, 1990:277-290.
- [10] Chung F L, Lee T. Fuzzy competitive learning[J]. Neural Networks, 1994,7(3):539-551.
- [11] Jacquin A E. Image coding based on a fractal theory of iterated contractive image transformations[J]. IEEE Transactions on IP, 1992,2(1):18-30.
- [12] Wohlberg B, Jager G. A review of the fractal image coding literature[J]. IEEE Transactions on IP, 1999,8(12):1716-1729.
- [13] 魏海. 基于小波的压缩域图像检索技术的初步研究[D]. 北京: 北京工业大学, 2000.
- [14] Shapiro M. Embedded image coding using zerotrees of wavelet coefficients[J]. IEEE Transactions on SP, 1993, 41:3445-3462.
- [15] 章毓晋. 图像工程[M]. 第二版. 北京: 清华大学出版社, 2003.
- [16] Arps R B, Truong T K. Comparison of international standards for lossless still image compression[C] // Proceedings of IEEE, 1994,82(6):889-899.
- [17] 刘伟华, 刘立柱, 张沛. JBIG 编码方式研究[J]. 信息工程大学学报, 2008, 9(1): 47-49.
- [18] 边文奎, 何丕廉, 等. 二值图像压缩标准——JBIG2[J]. 计算机工程与应用, 2002, 13:103-105.
- [19] 倪林, 苗原. 一种 JPEG2000 压缩域的图像检索方法[J]. 电子与信息学报, 2005, 27(3):474-477.
- [20] 李晓华. 小波压缩域图像检索技术的初步研究[D]. 北京: 北京工业大学, 2004.
- [21] 尹平, 汪宇飞, 胡迎新. JPEG2000 格式图像压缩算法的研究[J]. 计算机工程, 2007, 33(5):205-207.
- [22] 黄祥林. 基于压缩域的图像检索技术的初步研究[D]. 北京: 北京工业大学, 2001.

- [23] 向辉, 石教英. 压缩域多媒体数据处理技术研究[J]. 中国图像图形学报, 1999, 4(7):539-543.
- [24] 李晓华, 沈兰荪. 基于压缩域的图像检索技术[J]. 计算机学报, 2003, 26(9):1051-1059.
- [25] 黄翔宇, 章毓晋. 基于压缩域的图像检索技术研究进展[J]. 中国图像图形学报, 2003, 8(A): 499-508.
- [26] Idris F, Panchanatian S. Image indexing using vector quantization[C] // Proceedings of SPIE:Storage and Retrieval for Image and Video Databases, 1995, 2420:373-380.
- [27] Idris F, Panchanatian S. Storage and retrieval of compressed images[J]. IEEE Transactions on Consumer Electronics, 1995, 41(8):937-941.
- [28] Barbas J S, Wolk S I. Efficient organization of large ship radar databases using wavelets and structured vector quantization[C] // Proceedings of Asilomer Conference on Signals, Systems and Computers, 1993, 1:491-498.
- [29] Vellaikal A, Kuo C-C, Dao S. Content-based retrieval of remote sensed images using vector quantization[C] // Proceedings of SPIE:Visual Information Processing, 1995, 2488:178-189.
- [30] 魏海, 沈兰荪. 基于分类矢量量化的图像压缩和检索算法[J]. 电子学报, 2001, 29(7):933-936.
- [31] Eftekhari-Moghadam A M, Shanbehzadeh J, Mahmoudi F, et al. Image retrieval based on index compressed vector quantization[J]. Pattern Recognition, 2003, 36:2635-2647.
- [32] 邹彬, 潘志斌, 胡森. 基于局部投影与块 LBP 特征的图像检索[J]. 中国图像图形学报, 2012, 17(6):671-677.
- [33] Ramamurthi B, Gersho A. Classified vector quantization of image[J]. IEEE Transactions on Communications, 1986, 34(11):1105-1115.
- [34] Mandelbrot B. The fractal geometry of nature[M]. San Francisco, 1982.
- [35] 魏海, 沈兰荪. 一种基于小波分析的迭代分形编码方法[J]. 电路与系统学报, 1998, 3(4): 82-85.
- [36] Sloan A D. Retrieving database contents by image recognition:New fractal power[J]. Advanced Imaging, 1994, 9(5):26-30.
- [37] Zhang A, Cheng B, Acharya R S. Approach to query by texture in image databases system[C] // Proceedings of SPIE:Digital Image Storage and Archiving Systems, 1995, 2606:338-349.
- [38] Zhang A, Cheng B, Acharya R S, et al. Comparison of wavelet transform and fractal coding in texture-based image retrieval[C] // Proceedings of SPIE:Visual Data Exploration and Analysis III, 1996, 2666:116-125.

- [39] Ida T, Sambonsugi Y. Image segmentation using fractal coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1995,5(12):567-570.
- [40] 王志勇, 池哲儒, 余英林. 分形编码在图像检索中的应用[J]. 电子学报, 2000, 28(6):19-23.
- [41] 魏海, 沈兰荪. 反对称双正交小波应用于多尺度边缘提取的研究[J]. 电子学报, 2002, 30(3):313-316.
- [42] 杨守义, 罗伟雄. 一种分形域基于内容的图像检索方法[J]. 电子与信息学报, 2003, 25(3):419-423.
- [43] 陈添丁, 刘济林, 楼伟进. 基于分形编码拓扑特性的图像检索[J]. 中国图像图形学报, 2004, 9(1):56-61.
- [44] 张梁斌, 奚李峰. 一种基于熵及分形编码的图像检索方法[J]. 计算机工程与应用, 2008, 44(19):203-205.
- [45] Stone H S, Li C S. Image matching by means of intensity and texture matching in the Fourier domain[C] // Proceedings of SPIE, 1996, 2670:337-349.
- [46] Augustejin M L, Clemens E, Shaw K A. Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier[J]. IEEE Transactions on Geoscience Remote Sensing, 1995, 33(5):616-626.
- [47] 王文惠, 孟兵, 万建伟, 等. 利用不变量进行基于内容的图像检索[J]. 电子学报, 2002, 30(7):949-951.
- [48] 王斌. 一种基于多尺度拱高形状描述的图像检索方法[J]. 电子学报, 2013, 41(9): 1821-1825.
- [49] Celentano A, Lecce V D. A FFT based technique for image signature generation[C] // Proceedings of SPIE: Storage and Retrieval for Image and Video Databases V, 1997, 3022:457-466.
- [50] 沈兰荪, 魏海, 黄祥林. 压缩域图像处理技术研究[J]. 北京工业大学学报, 2000, 26(3):24-32.
- [51] Smith J R, Chang S F. Transform features for texture classification and discrimination in large image databases[C] // Proceedings of IEEE International Conference on Image Processing, 1994, 3:407-411.
- [52] Reeves R, Kubik K, Osberger W. Texture characterization of compressed aerial images using DCT coefficients[C] // Proceedings of SPIE: Storage and Retrieval for Image and Video Databases, 1997, 3022:398-407.
- [53] Bae H J, Jung S H. Fast texture based on DCT[C] // Proceedings of ICICS, 1997: 1065-1068.
- [54] Sim D G, Kim H K. Fast texture description and retrieval of DCT-based compressed images[J]. Electronics Letters, 2001, 37(1):18-19.

- [55] Furht B, Saksobhavit P. A Fast Content-Based Video and Image Retrieval Technique over Communication Channels[C] // Proceedings of SPIE Symposium on Multimedia Storage and Archiving Systems. Boston, MA, 1998.
- [56] 黄祥林, 宋磊, 沈兰荪. 基于 DCT 压缩域的图像检索方法[J]. 电子学报, 2002, 30(12):1786-1789.
- [57] 黄祥林, 沈兰荪. 一种具有旋转不变性的压缩域纹理图像分类方法[J]. 电子与信息学报, 2002, 24(11):1190-1196.
- [58] Fan Y, Wang R S. An image retrieval method using DCT feature[J]. Computer Science & Technology, 2002, 17(6):865-868.
- [59] Feng G C, Jiang J. JPEG compressed image retrieval via statistics features[J]. Pattern Recognition, 2003, 36(4):977-985.
- [60] Climer S, Bhatia S. Image database indexing using JPEG coefficients[J]. Pattern Recognition, 2002, 35:2479-2488.
- [61] Lay J A, Ling G. Image retrieval based on energy histograms of the low frequency DCT coefficients[J]. ICASSP, 1999, 6:3009-3012.
- [62] 许相莉, 张利彪, 于哲舟, 等. 基于 DCT 和 SVD 的图像检索算法[J]. 吉林大学学报, 2008, 46(6):1125-1130.
- [63] 王剑峰, 赵晓容, 李明科. 基于数字特征直方图的图像检索算法[J]. 重庆邮电大学学报, 2013, 25(5):700-704.
- [64] Abdelmalek A A, Hershey J E. Feature cueing in the discrete cosine domain[J]. Journal of Electronic Imaging, 1994, 3(1):71-80.
- [65] Shen B, Sethi I K. Direct feature extraction from compressed images[C] // Proceedings of SPIE:Storage and Retrieval for Image and Video Databases IV, 1996, 2670:404-414.
- [66] Lee S-W, Kim Y-N, Choi S W. Fast scene change detection using direct feature extraction from MPEG compressed videos[J]. IEEE Transactions on Multimedia, 2000, 2(4):240-254.
- [67] 黄祥林, 沈兰荪. 基于 DCT 压缩域的纹理图像分类[J]. 电子与信息学报, 2002, 24(2):216-221.
- [68] 张问银, 曾振柄, 孙星明. 基于 A/D 能量直方图的 JPEG 图像检索[J]. 计算工程, 2004, 16(30):21-22.
- [69] 李鹏杰, 杨树元. DCT 域中 MPEG7 主色描述符的提取[J]. 电子与信息学报, 2004, 26(11):1693-1699.
- [70] Shneier Michale, Mohamed Abdel-Mottaleb. Exploiting the JPEG compression scheme for image retrieval[J]. IEEE Transactions on PAMI, 1996, 18(8):849-853.

- [71] Yu Hong Heather. Visual image retrieval on compressed domain with Q-distance[C] // IEEE International Conference on Computational Intelligence and Multimedia Applications, 1999:1013-1016.
- [72] Chang C C, Chuang J C, Hu Y S. Retrieval digital images from a JPEG compressed image database[J]. Image and Vision Computing, 2004, 22:471-484.
- [73] Zhong Y, Zhang H J, et al. Automatic caption localization in compressed video[J]. IEEE Transactions on PAMI, 2000, 22(4):385-392.
- [74] 黄祥林, 沈兰荪. 基于 DCT 压缩域的图像字符定位[J]. 中国图像图形学报, 2002, 7(A), (1):22-26.
- [75] Mallat S. A theory for multiresolution signal decomposition: the wavelet representation[J]. IEEE Transactions on PAMI, 1989, 11(7):674-693.
- [76] Mandal M K, Aboulnasr T, Panchanathan S. Fast wavelet histogram techniques for image indexing[J]. Journal of Computer Vision and Image Understanding, 1999, 75(1):99-110.
- [77] Jacobs C E, Findelstein A, Salesin D H. Fast multiresolution image querying[C] // ACM International Conference on Computer Graphics and Interactive Techniques, 1995:277-286.
- [78] Wang J Z, Wiederhold G, Firschein O, et al. Wavelet-based image indexing techniques with partial sketch retrieval capability[C] // International Forum on Research and Technology Advances in Digital Libraries, 1997:13-24.
- [79] 闫允一, 姜帅, 郭宝龙. 结合稳定兴趣点和 Gabor 小波的图像检索[J]. 西安电子科技大学学报, 2014, 41(5).
- [80] Chang T, Kuo C-C J. Texture analysis and classification with tree-structured wavelet transform[J]. IEEE Transactions on Image Processing, 1993, 2(4):429-441.
- [81] Lee M-C, Pun C-M. Texture classification using dominant wavelet packet energy features[C] // Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, 2000:301-304.
- [82] Albanesi M G, Giacane A. Fast retrieval on compressed images for internet applications[C] // Proceedings of the 5th IEEE International Workshop on Computer Architectures for Machine Perception, 2000:136-141.
- [83] Ma W Y, Manjunath B S. A comparison of wavelet transform features for texture image annotation[J]. ICIP, 1995:256-259.
- [84] Chang S-F. Compressed domain techniques for image/video indexing and manipulation[J]. ICIP, 1995:314-317.
- [85] Mandal M K, Aboulnesr T, Panchanatian S. Image indexing using moments and wavelets[J]. IEEE Transactions on Consumer Electronics, 1996, 42(3):557-565.

- [86] Mandal M K. Wavelet based coding and indexing of images and video[D]. Canada: University of Ottawa, Canada, 1998.
- [87] Chen J L, Kundu A. Rotation and gray scale invariant texture identification using wavelet decomposition and hidden Markov model[J]. IEEE Transactions on PAMI, 1994, 16(2):208-214.
- [88] Sun M, Sclabassi R J. Symmetric wavelet edge detector of the minimum length[J]. ICIP, 1995:177-180.
- [89] 魏海, 沈兰荪. 小波变换域内基于方向梯度相角直方图的图像检索算法[J]. 电路与系统学报, 2001, 6(2):20-24.
- [90] 魏海, 沈兰荪, 李晓华. 基于迭代分形的图像压缩和检索方法[J]. 中国图像图形学报, 2002, 7(A)11:1198-1203.
- [91] 姚玉荣, 章毓晋. 利用小波和矩进行基于形状的图像检索[J]. 中国图像图形学报, 2000, 5(3):206-210.
- [92] 周燕, 曾凡智, 卢炎生, 等. 基于压缩感知的图像检索方法研究[J]. 中山大学学报, 2014, 33(1):57-66.
- [93] Wilson B, Bayoumi M A. Compressed domain classification of texture images[C] // Proceedings of the 5th IEEE International Workshop on Computer Architectures for Machine Perception, 2000:347-355.
- [94] 牛蕾, 倪林. 基于 ROI 的压缩域多谱段遥感图像的检索[J]. 中国图像图形学报, 2005, 10(10):1212-1217.
- [95] Qi F, Shen D, Quan L. Wavelet transform based rotation invariant feature extraction in object recognition[C] // Proceedings of International Symposium Information Theory&Its Application, 1994, 11:221-224.
- [96] Rashkovshiy P, Sadovnik L. Scale, rotation and shift invariant wavelet transforms[C] // Proceedings of SPIE:Optical Pattern Recognition, 1994, V2237:390-401.
- [97] Froment J, Mallat S. Second generation image coding and wavelet transform[C] // Proceedings of International Conference First World Congress of Nolinear Analysis, 1996:1923-1932.
- [98] Loupiaz E, Sebe N, Bres S, et al. Wavelet-based salient points for image retrieval[C] // Proceedings of IEEE International Conference on Image Processing, 2000:518-521.
- [99] Sebe N, Lew M S, Tian Q, et al. Color indexing using wavelet-based salient points[C] // IEEE Workshop on Content-based Access of Image and Video Libraries, 2000: 15-19.
- [100] Liu C, Mandal M. Image indexing in the JPEG2000 framework[C] // Proceedings of SPIE:Internet Multimedia Management System, 2000, 4210:272-280.
- [101] Liu C, Mandal M. Fast image indexing based on JPEG2000 packet header[C] //

- Proceedings of 3rd International Workshop on Multimedia Information Retrieval, 2001.
- [102] Bhalod J, Fahmy G F, Panchanathan S. Region based indexing in the JPEG-2000 framework[C] // Proceedings of SPIE: Internet Multimedia Management Systems II, 2001, 4519:91-96.
- [103] Xiong Ziyu, Huang Thomas S. Block-based, Memory-efficient JPEG2000 image indexing in compressed-domain[C] // Fifth IEEE Southwest Symposium on Image Analysis and Interpretation, 2002:92-95.
- [104] Lin Ni. A novel image retrieval scheme in JPEG 2000 compressed domain based on tree distance[C] // The Fifth International Conference on Information and Communications Security, 2003:1591-1594.
- [105] 章毓晋. 图像处理和分析基础[M]. 北京: 高等教育出版社, 2002.
- [106] Pentland A, Picard R W, Sclaroff S. Photobook: tools for content-based manipulation of image database[C] // Proceedings of SPIE:Storage and Retrieval for Image and Video Databases, 1994, 2185:34-37.
- [107] Swets D L, Weng J. Using discriminant eigenfeature for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8):831-836.
- [108] 曹奎, 冯玉才, 王元珍. 一种基于颜色的图像表示及全局相似检索技术[J]. 计算机研究与发展, 2001, 38(9):1121-1126.
- [109] 杨琼, 丁晓青. 对称主分量分析及其在人脸识别中的应用[J]. 计算机学报, 2003, 26(9):1146-1151.
- [110] 万其明, 汪闽, 张星月, 等. 基于二叉树分解与多特征直方图匹配的高分辨遥感图像检索[J]. 地理信息科学学报, 2010, 12(2):275-281.
- [111] Idris F, Panchanathan S. Image indexing using wavelet vector quantization[C] // Proceedings of SPIE: Digital Image Storage Archiving Systems, 1995, 2606:269-275.
- [112] Swanson M D, Hosur S, Tewfik A H. Image coding for content-based retrieval[J]. SPIE, 1996, 2727:4-151.
- [113] Podilchuk C, Zhang X. Face recognition using DCT-based feature vector[J]. ICASSP, 1996, 4:2144-2147.
- [114] Swanson W M, Tewfik A H. Fast progressively refined image retrieval[J]. Journal of Electronic Imaging, 1998, 7(3):443-452.
- [115] 赵珊, 周利华. DCT 压缩域中基于纹理和形状的图像检索算法[J]. 西安电子科技大学学报, 2007, 34(3):402-408.
- [116] 沈恩华, 邱志诚, 孟欣, 等. 脑电图复杂度分析中的粗粒化问题: 量化对复杂度计算的影响[J]. 生物物理学报, 2000, 16(4):707-710.

- [117] 陈芳, 顾凡及, 徐京华, 等. 一种新的人脑信息传输复杂性的研究[J]. 生物物理学报, 1998, 14(3):508-512.
- [118] 沈恩华. 脑电的复杂度分析[D]. 上海: 复旦大学, 2004.
- [119] 蔡志杰, 顾凡及, 沈恩华. Co 复杂度的数学基础[J]. 应用数学与力学, 2005, 26(9): 1188-1196.
- [120] 赵珊, 周利华. DCT 压缩域的图像检索[J]. 北京邮电大学学报, 2007, 30(6):107-110.
- [121] Zhao S. Image retrieval based on edge in DCT compressed domain[J]. ITESS, 2008, VI:1104-1108.
- [122] 赵珊, 刘静. DCT 压缩域的图像检索技术[J]. 北京邮电大学学报, 2008, 31(5):5-8.



## 视觉注意计算模型

视觉注意是利用获取的视觉信息进行注意选择的一种心理现象，可以使视觉系统在处理海量信息时，把有限的资源优先分配给少数几个显著的区域，而对同时进入它们视野的不太重要的信息则视而不见，保证了视觉系统对获取的信息进行有选择性的、实时的处理。本章在介绍人类视觉系统和视觉系统理论的基础上，引入了 3 种视觉注意计算模型。

### 4.1 概 述

#### 4.1.1 人类视觉系统

视觉系统是神经系统的一个组成部分，它使生物体具有了视知觉能力。当环境中的可见光信息进入视觉系统后，生物体就会产生相应的刺激响应，从而实现对周围世界的感知。这里所说的可见光是一个广义的概念，不同物种所能感知的可见光也不相同。例如，人类只能看到可见光（狭义概念，指通常意义下的可见光），而有些物种可以看到紫外部分，另一些则可以看到红外部分。视觉系统具有将外部世界的二维投射重构为三维世界的能力。

人类的视觉系统包括眼（特别是视网膜）、视神经、视交叉、视束、外侧膝状体、视辐射、视皮层、中颞（Middle Temporal, MT）区等，其结构图如图 4.1 所示<sup>[1]</sup>。

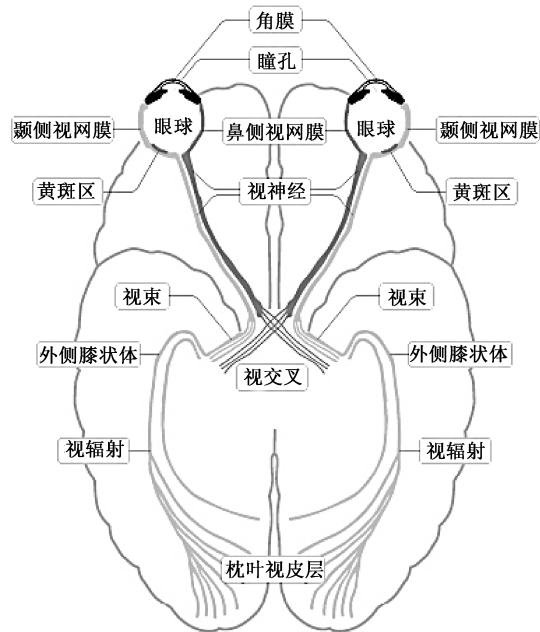


图 4.1 人类视觉系统的结构图

视网膜的结构图如图 4.2 所示。在视网膜中包含大量的光感受器细胞，这些细胞含有一种称为视蛋白的蛋白分子。视蛋白的作用就是接收外界来的光子，然后将刺激信号通过信号传导通路传递给光感受器细胞，使其发生超极化现象，产生相应的触发动作。人类的光感受器细胞有两种：视杆细胞和视锥细胞。视杆细胞和视锥细胞分别负责不同的工作。当光线很弱时，分布在视网膜周边部分的视杆细胞是工作的主力军；而在正常光强条件下，处于视网膜中心（由于视网膜贴在眼球的后方，中间部分是一个凹形的区域，因而也称作中央凹）的视锥细胞开始工作，负责辨别颜色及接收其他

视觉信息。因此，视锥细胞的作用比视杆细胞的作用更大些。不同的视锥细胞吸收不同波长的光子，由此可以将视锥细胞分为 3 类：短/蓝视锥细胞（吸收短波长的蓝光）、中/绿视锥细胞（吸收中波长的绿光）和长/红视锥细胞（吸收长波长的红光）。

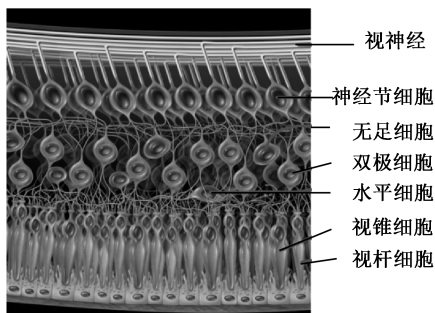


图 4.2 视网膜的结构图

视杆细胞和视锥细胞的突触直接与双极细胞相连，而双极细胞的突触又与节细胞相连，最后由节细胞将动作电位传递到视皮层。在人类视觉系统中，大约有 1 亿 3 千万个光感受器，接收到光信号后，通过大约 120 万个节细胞轴突将信息传递到大脑，大量的视觉处理过程就是在这个庞大的视神经网络中完成的。

其中包括形成双极细胞及节细胞的中心-周边感受野，以及从光感受器到双极细胞的信息汇聚和发散。神经网络的横向信息传递由水平细胞和无足细胞完成，以形成更加复杂的感受野，如对深度敏感而对颜色不敏感的感受野或者对颜色敏感而对外形和深度不敏感的感受野。

节细胞共有 5 种：①M 细胞有大的中心-周边感受野，对深度敏感，对颜色不敏感，能对刺激迅速产生响应；②P 细胞有小的中心-周边感受野，对颜色和形状敏感；③K 细胞只有非常大的中心感受野，对颜色敏感，对外形和深度不敏感；④第 4 种节细胞具有内在的光敏性；⑤最后一种节细胞主要用于眼动。

最后，视觉信息沿着视神经传递到视皮层，实现了由眼传递到大脑的全过程。视神经中有大约 90% 的轴突到达丘脑的感觉中继核团——外侧膝状体核（Lateral Geniculate Nucleus, LGN）。人类和其他起源于 catarrhiniens [包括 cercopithecidae（猕猴科）和 apes（猿、猩猩等）] 的灵长类动物的 LGN 共有 6 层，有点像现在大家常用的信用卡的结构，厚度是普通信用卡的 3 倍左右，但形状上不是平的，而是弯曲成椭球面。两只眼的信息分别由对侧的第 1、4、6 层和第 2、3、5 层接收。第 1 层包含 M（大）细胞，与深度或运动视觉有关，第 4、6 层与 P（小）细胞（颜色与边界）形成连接，第 2、3、5 层类似。还有一些更小的细胞夹在 6 层中间，负责接收来自于视网膜 K 细胞（颜色）的信息。LGN 神经元将视觉信息传递到初级视皮层（V1），初级视皮层位于大脑的后部枕叶区邻近距状沟。

## 4.1.2 视觉系统理论

### 1. 有效编码假说

半个世纪前，人类对自身神经系统的认识还不是很多，而且也没有现在的信息理论，因而无法从本质上理解人类视觉系统。但是复杂的外部环境和有限的神经元数量之间的矛盾明显地摆在那里，研究者不得不利用当时有限的理论知识，去设想人类视觉系统的工作原理，由此便产生了有效编码假说。但在从提出开始的几十年里，并未有太多进展，直到最近十多年，随着神经学方面的深入研究，以及信息科学的不断发展，这一假说才得到了迅猛发展。

面对多姿多彩的外界世界，人类视觉系统能够从容应对，给人类以美好的视觉体验，而不发生信息爆炸，这得益于长期的进化。这种机理到底是什么？研究者认为感知系统的信息处理过程肯定受到外界信号统计模型的影响。Attneave<sup>[2]</sup>认为，感知就是对外来的输入信号进行有效表示。Barlow<sup>[3]</sup>进一步从理论上研究，根据当时信息处理理论中最有效的香农信息论，提出了有效编码假说，认为视觉系统之所以能够面对外界的大量刺激而不会爆炸，一个重要的约束就是信息（或者编码）的有效性。纷繁复杂的外部刺激存在大量的冗余，大脑的神经元只有有效地去除这些冗余，才能利用较

少的资源尽可能有效地表达更多的信息，从而自适应外界环境。

在有效编码假说的指导下，研究人员进行了大量卓越的工作。Vinje 和 Gallant<sup>[4]</sup>在神经生理学实验中发现，视皮层细胞对外界刺激的响应满足稀疏分布。Olshansen 和 Field 等<sup>[5]</sup>认为神经系统处理信息过程中产生不变特征，并通过线性变换（当然，真实的大脑远不止这么简单）模拟神经元细胞之间的相互作用验证了这一点。Olshansen<sup>[6]</sup>又提出动态开关网络，使用归一化方法，提取物体发生平移和尺寸变化时的不变特征。Fukushima<sup>[1]</sup>提出神经认知机，其本质上是一个层次化的神经网络。因为有多层神经元，因此神经认知机能识别变形的模式，但参数也较多，由此带来的困难也随着增加。Field<sup>[7]</sup>和 Daugman<sup>[8]</sup>分别证明了自然图像在高阶统计特性上的非高斯性，即面对外界信号，大部分神经元保持沉默，只有少量神经元反响强烈。依据这种特性——稀疏性（sparseness），Olshausen<sup>[5]</sup>设计了稀疏编码（sparse coding）模型，用来提取自然图像的不变特征，取得了很好的实验结果。

通过对有效编码假说的分析，可以看出，人类神经系统的感知过程可以认为是对外界刺激的一种有效编码，其中提取不变特征是关键。广义上的特征提取是指一种变换或一种编码表达。

## 2. Marr 视觉理论

在认知科学领域中，“视知觉从哪里开始”是一个本源性问题，也就是说视知觉

过程到底是从局部开始还是从整体开始，目前没有定论。Marr<sup>[9]</sup>在 20 世纪 70 年代提出的特征分析理论认为，视知觉过程是由局部性质到整体性质进行的，视觉系统首先感知到的是局部几何特征，而后才会感知到大范围的拓扑特征。而由 Navon 提出的整体优先理论认为视知觉过程是由大范围整体拓扑性质到局部性质进行的。

Marr 认为，视觉就是要对外部世界的图像构成有效的符号描述，它的核心问题是要从图像的结构推导出外部世界的结构。视觉从图像开始，经过 3 级结构“要素图—2.5 维图—三维模型表象”，对其进行一系列的处理和转换，最后达到对外部现实世界的认知，具体过程如图 4.3 所示。

（1）要素图：主要指图像中强度变化剧烈处的位置及其几何分布和组织结构，其中用到的基元包括斑点、端点、边缘片断、有效线段、线段组、曲线组织、边界等。由图像生成要素图是视觉过程的第一个阶段，也称为早期视觉，其目的是要把原始二维图像中的重要信息更清楚地表示出来。

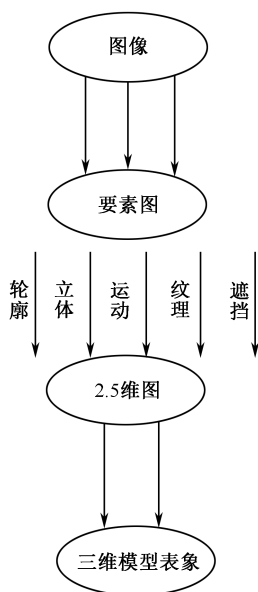


图 4.3 视觉系统的 3 个表象层次

(2) 2.5 维图：是指在以观察者为中心的坐标系中，可见表面的法线方向、大致的深度及它们的不连续轮廓等，其中用到的基元包括可见表面上各点的法线方向、各点离观察者的距离（深度）、深度上的不连续点、表面法线方向上的不连续点等。之所以称为 2.5 维图，是因为其中包含了二维图中没有的深度信息，但还不是真正的三维表示。由要素图到 2.5 维图是视觉过程的第二个阶段，也称为中期视觉。按照 Marr 的理论，这一阶段由一系列相对独立的处理模块组成，包括运动、由表面明暗恢复形状、由表面轮廓线恢复形状、由表面纹理恢复形状等，其作用是揭示一个图像的表面特征。

(3) 三维模型表象：指的是在以物体为中心的坐标系中，用含有体积基元（即表示形状所占体积的基元）和面积基元的模块化分层次表象，描述形状和形状的空间组织形式，其表征包括容积、大小和形状等。由 2.5 维图获得物体的三维模型表象是视觉过程的第三个阶段，也称为后期视觉，其作用是得到物体的一种独特的描述。

### 3. Treisman 特征整合理论

注意的特征整合理论（Feature Integration Theory, FIT）主要探讨视觉早期加工的问题，由 Treisman、Sykes 和 Gelade 于 1980 年提出<sup>[10]</sup>。具体的视觉加工过程分为两个阶段（如图 4.4 所示）。

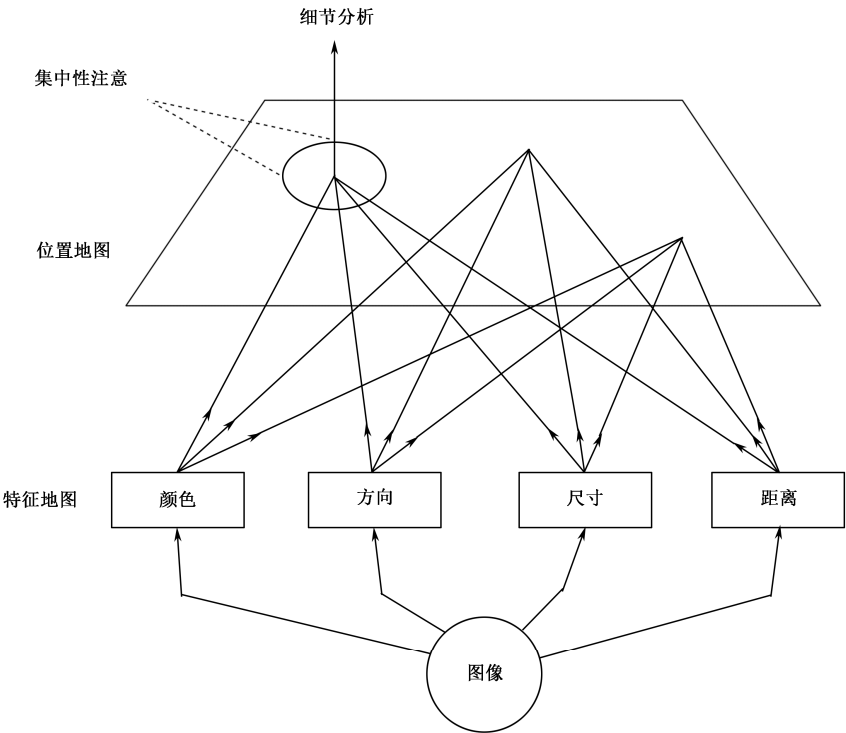


图 4.4 注意的特征整合理论

(1) 特征登记阶段 (又称为前注意阶段)。在这一阶段, 人几乎意识不到注意的发生。视觉系统提取颜色、尺寸、方向、反差、倾斜性、曲率和线段端点等特征, 并进行平行的、自动化的加工。这些特征处于自由漂浮状态 (free-floating state), 并被独立编码, 得到特征地图 (feature map)。

(2) 特征整合阶段 (又称为物体知觉阶段)。在这一阶段, 知觉系统对特征进行定位, 得到位置地图 (map of locations), 以此把彼此分开的特征正确联系起来, 形成对某一物体的表征。然后, 集中性注意就像胶水一样, 把原始的、彼此分开的特征整合为一个单一的物体。这一系列加工过程是一种非自动化的、序列的处理过程, 因而比前一阶段要慢一些。由于需要努力, 当注意超负荷或人们分心时, 特别是对注意的要求很高时, 就可能会把外界刺激的特征不恰当地组合在一起, 从而造成错觉现象。

#### 4. 视觉注意计算模型理论

虽然目前人们对于注意的生理机制还不是完全清楚, 但研究者从计算的角度对其进行了研究, 力图用合适的数学模型模拟视觉注意能力。研究者普遍认为, 视觉注意机制体现在两个方面: 一方面是由视觉信息的强烈刺激而产生的, 常被称为由数据驱动的注意或自底向上的注意; 另一方面是由人类为了执行某个任务而主动发起的, 常被称为由任务驱动的注意或自顶向下的注意。这两种注意机制实质上并非完全独立, 而是互相交织在一起, 相辅相成, 互相促进的。

Itti 视觉注意计算模型<sup>[11]</sup>是一种数据驱动的注意模型, 主要基于 Treisman 特征整合理论。首先, 从输入图像中提取多方面的特征, 如颜色、朝向、亮度、运动等, 运用一些数学工具对其进行处理, 得到各个特征维上的显著图; 然后, 对这些显著图进行分析、融合得到兴趣图; 最后, 通过一定的竞争机制, 从兴趣图中的多个待注意的候选目标中选出唯一的注意目标, 具体流程图如图 4.5 所示。

从图像中提取亮度、颜色、朝向等初级视觉特征, 在模型中采用线性滤波器组对图像滤波的方法。由于高斯滤波器具有归一性、对称性、单峰性和对下一层的等贡献性等优点<sup>[12]</sup>, 因而常被用在初级视觉特征的提取过程中。此外, 一些由高斯滤波器经过各种变换得到的滤波器也被经常用到, 其中最为有用的是 Gabor 滤波器。Gabor 滤波器是对高斯滤波器函数 (简称高斯函数) 的正弦或余弦调制, 能够很好地提取自然图像中目标的朝向特征。在 Itti 视觉注意计算模型中, 采用了高斯差分滤波器实现对图像的多尺度采样。具体来说, 就是将不同尺度的滤波器组合在一起, 形成金字塔结构 (也就是常说的高斯金字塔), 然后对输入图像逐级进行低通滤波和子采样, 从而得到同一幅图像的不同分辨率表示。其中, 金字塔的最底层是原始图像, 由此往上, 对每一层图像先滤波再采样, 就获得了新的一层图像, 新图像的大小在水平和垂直方向都是原图像的二分之一。

在 Itti 视觉注意计算模型中, 金字塔共分为 9 层, 原始图像为第 0 层, 在水平和垂直方向分别以 2 为因子对图像分辨率进行递减, 得到第 1、2、……、8 层, 最小的

一层分辨率为原来的 1/256。

用一个 5×5 的高斯模板对金字塔的每一层图像都进行低通滤波。

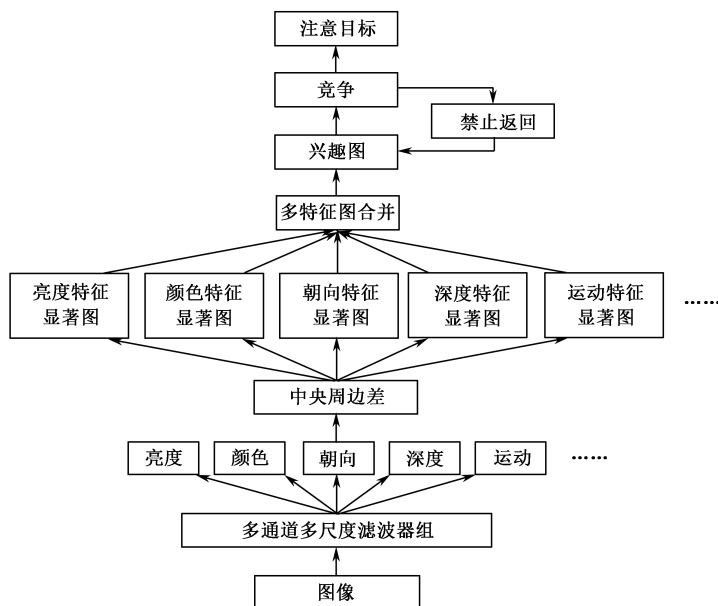


图 4.5 Itti 视觉注意计算模型的流程图

高斯滤波器函数的计算公式为式 (4-1)，其宽度（决定着平滑程度）是由参数  $\sigma$  表征的，而且  $\sigma$  和平滑程度的关系是非常简单的。 $\sigma$  越大，高斯滤波器的频带就越宽，平滑程度就越好；反之， $\sigma$  越小，高斯滤波器的频带就越窄，平滑程度就越差。因此，可以通过调节平滑程度参数  $\sigma$ ，使滤波效果达到最佳，即既不会由于过平滑而使得图像特征过分模糊，也不会由于没有对图像中的噪声和细纹理进行充分的平滑而导致出现过多的不希望突变量。

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (4-1)$$

选择高斯滤波器，还有另外一个原因，那就是高斯函数的可分离性。二维高斯函数卷积可以分成两个一维高斯函数来进行，而在计算量上却只是线性增长而非平方增长。

由于人的视网膜分为中央凹（fovea）和外围（periphery）两个部分，中央凹的分辨率高，而外围的分辨率低，绝大部分信息处理是在中央凹完成的，因而这种特征被称为非均匀采样<sup>[13]</sup>。在 Itti 提出的高斯金字塔模型中，非均匀采样体现在不同的采样层次和中央周边差的采样方式上。

众所周知，吸引视觉注意的是最“与众不同”的部分，也就是说，它与周围其他部分相比具有更高的显著性。因此，特征提取的输出可以设计为特征的对比度，即视

点中央与周围部分的差值。在实际计算中,特征对比度转化为图像特征图在不同尺度下的差值。具体做法是对粗尺度下的特征图进行插值,转变为与细尺度同样大小的特征图,然后再将这两幅特征图进行点对点减法。由于细尺度能发现高频部分,检测的是小的图像区域,代表中央区域,而粗尺度能发现低频部分,检测的是大的图像区域,代表周边区域,因此这种方法称为中央周边差采样方式。

具体地,假设输入图像的红色、绿色、蓝色 3 个颜色通道分别用  $r$ 、 $g$ 、 $b$  来表示,则亮度通道  $I$  可以用式 (4-2) 计算得到。

$$I = (r + g + b) / 3 \quad (4-2)$$

按照前面提到的思路,对亮度  $I$  进行多层高斯滤波,就得到了亮度高斯金字塔  $I(\sigma)$ , 其中  $\sigma$  为金字塔的层次,取值为  $\sigma \in [0, 8]$ 。

$$I(\sigma) = I * G(x, y, \sigma) \quad (4-3)$$

其中,  $G(x, y, \sigma)$  为高斯滤波器函数。

根据对比色学说,构建颜色金字塔时,分别计算红色、绿色、蓝色和黄色 4 个颜色通道,并最后合并为红/绿和蓝/黄两个对比颜色通道。红色、绿色、蓝色和黄色 4 个颜色通道记为  $R$ 、 $G$ 、 $B$ 、 $Y$ , 计算公式如式 (4-4) 至式 (4-7) 所示。

$$R = r - (g + b) / 2 \quad (4-4)$$

$$G = g - (r + b) / 2 \quad (4-5)$$

$$B = b - (r + g) / 2 \quad (4-6)$$

$$Y = r + g - 2(|r - g + 2|) \quad (4-7)$$

与亮度通道类似,对红色、绿色、蓝色和黄色 4 个颜色通道  $R$ 、 $G$ 、 $B$ 、 $Y$  进行多层高斯滤波,就可以得到颜色高斯金字塔  $R(\sigma)$ 、 $G(\sigma)$ 、 $B(\sigma)$ 、 $Y(\sigma)$ , 其中  $\sigma$  为金字塔的层次,取值为  $\sigma \in [0, 8]$ 。

$$R(\sigma) = R * G(x, y, \sigma) \quad (4-8)$$

$$G(\sigma) = G * G(x, y, \sigma) \quad (4-9)$$

$$B(\sigma) = B * G(x, y, \sigma) \quad (4-10)$$

$$Y(\sigma) = Y * G(x, y, \sigma) \quad (4-11)$$

正如前边提到的, Gabor 滤波器在提取目标的朝向特征方面有很大的优势,因而,在 Itti 视觉注意计算模型中,采用 Gabor 滤波器来提取朝向特征。Gabor 滤波器是对高斯函数的正弦或余弦调制。二维 Gabor 函数可以表示为

$$h(x, y) = g(x', y') [\cos(2\pi f_0 x') + j \sin(2\pi f_0 x')] \quad (4-12)$$

其中,

$$g(x, y) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \quad (4-13)$$

$$x' = x \cos \theta + y \sin \theta \quad (4-14)$$

$$y' = -x \sin \theta + y \cos \theta \quad (4-15)$$



一般情况下,模型中只考虑水平、垂直、正反对角线4个方向,因此构造出的朝向通道有4个,记为 $O(\sigma, \theta)$ ,其中尺度 $\sigma \in [0, 8]$ ,方向 $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ 。

$$O(\sigma, \theta) = I * h(x, y) \quad (4-16)$$

这样,就得到1个亮度特征图组、2个颜色特征图组和4个朝向特征图组,共7组。每一组特征图组中有9层特征图。

对每一组特征图组执行中央周边差操作(记为 $\ominus$ ),就可以得到特征显著图。感受野的中央像素点位于金字塔的 $c$ 层,而周边像素点位于金字塔的 $s$ 层,其中 $s = c + \delta$ ,  $c \in \{2, 3, 4\}$ ,  $\delta \in \{3, 4\}$ 。因此,对每一类特征,都可以得到6个特征显著图,分别为2-5、2-6、3-6、3-7、4-7、4-8,如图4.6所示。最后,通过计算可以得到共42幅特征显著图,包括6幅亮度特征显著图、12幅颜色特征显著图和24幅朝向特征显著图。

具体地,亮度特征显著图组由式(4-17)计算得到,提取到的显著区域具有中央黑周边白或中央白周边黑的特点。

$$J(c, s) = |I(c) \ominus I(s)| \quad (4-17)$$

根据对比色学说,红色、绿色、蓝色和黄色4个颜色通道最终合并为红/绿和蓝/黄两个对比颜色通道。颜色特征显著图组的对应计算公式为

$$RG(c, s) = |[R(c) - G(c)] \ominus [R(s) - G(s)]| \quad (4-18)$$

$$BY(c, s) = |[B(c) - Y(c)] \ominus [B(s) - Y(s)]| \quad (4-19)$$

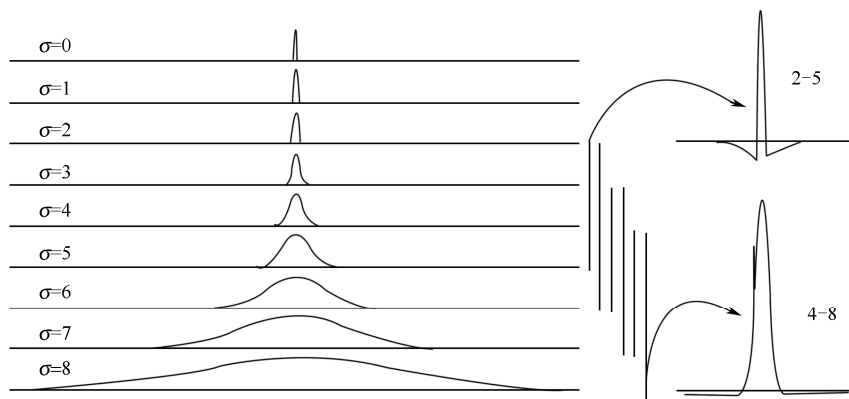


图4.6 中央周边差操作示意图<sup>[11]</sup>

同样,朝向特征显著图组按照 $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ 的取值共分为4组,每组6幅,计算公式为

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (4-20)$$

接下来,按照一定的策略将这些特征显著图合并成为一张兴趣图 $S$ 。多特征显著图合并采用的最简单的处理方式是直接相加。由于计算的不同特征通道的显著值变化范围不一样,所以需要先执行一个归一化操作 $N(\cdot)$ ,把各种特征值归一化到同一个范

围内, 如 0~1 的范围内。

$$S = \frac{1}{3} \left( N \left\{ \bigoplus_c \bigoplus_s N[J(c, s)] \right\} \right) + N \left( \bigoplus_c \bigoplus_s \left\{ N[RG(c, s)] + N[BY(c, s)] \right\} \right) + N \left\{ \sum_{\theta} \bigoplus_c \bigoplus_s N[O(c, s, \theta)] \right\} \quad (4-21)$$

这种直接相加的方法没有考虑不同特征的优先级, 容易引起互相抵消的情况。因而, 在后来的模型中, 引入了非线性的合并策略等方法, 以便避免不同的显著特征之间的冲突。

在这幅兴趣图中, 有很多显著区域, 这便是待注意目标, 模型采用最简单的胜者为王 (winner-take-all) 的竞争机制来确定注意焦点的位置, 使最显著的区域成为当前注意的焦点。

为了能使注意焦点在各个待注意目标之间转移, 使每个显著区域都有可能吸引注意焦点, 还需要加入另一个机制——禁止返回机制 (inhibition of return), 即一旦在一个区域已经搜索过或没有发现与目标有关的线索, 那么这个区域在以后的目标搜索中将不再被列入查找范围。

### 4.1.3 研究现状

自 20 世纪中期以来, 研究者对注意机制进行了大量研究, 从心理学实验总结出了一些有意义的模型。在此基础上, 研究者通过进一步对人类视觉系统的注意机制进行不懈探索, 尝试用数学和计算的方法, 构建了能够在一定程度上模拟人类视觉系统的选择性注意能力的视觉注意计算模型, 取得了长足的进步。本小节对现有的选择性注意理论和视觉注意计算模型的研究状况进行总结, 并分析其中的一些不足。

#### 1. 选择性注意理论方面的研究现状与问题分析

自 20 世纪中期以来, 研究者提出了很多选择性注意理论, 其中最有代表性的理论有过滤器模型<sup>[14]</sup>、衰减器模型<sup>[15]</sup>、特征整合理论<sup>[16]</sup>、创造性综合选择说<sup>[17]</sup>、后期选择模型<sup>[18]</sup>及资源限制说<sup>[19]</sup>等。

这些理论或学说可以在一定条件下对人的选择性注意给予合理的解释, 但也都存在一些缺陷, 对于其他模型提出的一些现象无法作出合理的解释。例如, 过滤器模型认为注意处于信息处理过程的早期阶段, 系统用一个过滤器来应对来自外界的大量信息, 放过一些信息, 而过滤其他一些信息。这一模型在双耳听音实验中得到了验证。但另外一些实验证明, 对被试者而言十分重要的信息 (如被试者的名字等) 则同样能够通过过滤器, 因此研究者又提出了衰减器模型, 认为对外界信息的处理不是按照“全或无”的形式被过滤的, 而是一个信息的衰减过程, 从而很好地解释了鸡尾酒会效应 (指在一个鸡尾酒会上, 正在聊天的人在十分嘈杂的环境中也能听到远处其他人谈论

的有关他的话)。特征整合理论则认为外界信息首先是被全部平行加工的,而在加工之后,才进行整合处理,进而产生注意。后期选择模型也持类似观点,认为注意是在信息处理过程的晚期进行的,当然支持这类理论的实验是在外界刺激负载比较小的情况下完成的,而在刺激负载比较大的情况下不能很好地解释。因而,研究者又提出创造性综合选择说,认为注意是人们对刺激知觉结果的积极、主动的预测能力。另外,人的注意还受到唤醒、心理资源、情绪、药物等因素的影响,资源限制说认为注意是中枢神经系统对于可用执行任务的一种资源分配能力。

这些理论或学说对人的注意能力是如何表现的进行了一些研究,但是所有这些模型都只研究了实现“选择信息”的各种可能的实现方法,而没有揭示注意能力生成的本质机理是什么。例如,过滤器模型没有解释清楚究竟依据怎样的准则来确定什么信息应当过滤,什么信息应当衰减,什么信息不应当衰减,如此等等。我们注意到,人类的注意能力,是人类在感知到外界信息之后,对于巨量信息所做出的第一步反应,因而探讨注意生成的机制有着非常重要的意义。

## 2. 视觉注意计算模型的研究现状分析(研究对象角度)

由于在认知科学领域中,关于人类视觉系统到底是先注意到局部特征而后才形成整体特征,还是先注意到整体特征后再分析其局部特征这一本源问题的理解存在分歧,因此视觉注意计算模型的研究方式也不尽相同。有的倾向于整体优先,并且强调注意物体的完整性,便形成了基于物体的视觉注意计算模型;有的则倾向于局部优先,并认为局部特征可以独立对待,而后再融合成统一的空间上的显著区域,这些思路大多属于基于空间的视觉注意计算模型。此外,在神经生理学的研究成果指导下,研究者通过模拟生物的两条视觉通路——“what”通路和“where”通路,提出了更具仿生性的基于视觉通路的视觉注意计算模型。

### 1) 基于空间的视觉注意计算模型

认同这种模型的研究者把视觉注意比喻成一个“探照灯”<sup>[16]</sup>,其关注点在于视觉注意焦点是不是空间中的某一个空间点,而不管是不是聚焦到某一个完整的物体上。这种思想的理论基础有两个:最主要的是特征综合理论(Feature Integration Theory, FIT)<sup>[16]</sup>,由 Treisman 等人首次提出;另外一个就是由 Koch 和 Ullman<sup>[20]</sup>提出的结构,这种结构的基础是人类的神经系统。FIT 认为人类视觉系统是以并行的方式对信息进行处理,也就是说,不同的神经元负责注意不同的特征(如颜色、朝向、运动等)。

在这两个理论的基础上,有些研究者提出了几个很有效的视觉注意计算模型,其中效果最突出的是 Koch 和 Ullman 提出的数学模型。后来,Itti 等人<sup>[21-27]</sup>首次实现了该模型,成为基于空间的视觉注意计算模型中的佼佼者,近十几年来,基于空间的视觉注意计算模型的研究基本上没有突破这一框架。他们从视觉分析角度出发,模拟人类视觉系统的视觉注意体系的功能,构建出的模型应用在图像处理、物体识别等方面

取得了比较好的效果<sup>[28-30]</sup>。

在 Itti 的最初模型中<sup>[27]</sup>，对于视觉信息的处理，采用了如下思路：①提取一些底层视觉特征，如亮度、颜色、朝向等；②计算各底层通道的高斯金字塔图像组；③用中央周边差操作对高斯金字塔图像组进行处理，得到特征显著图；④融合所有特征显著图形成一个总的兴趣图；⑤采用禁止返回机制来完成兴趣图上多个注意焦点间的转移。后来，Itti 对此模型进行了一些改进，如在特征通道上增加了深度特征<sup>[31,32]</sup>等、采取非线性的特征显著图的合并策略<sup>[24,33,34]</sup>及增加自顶向下的特征通道<sup>[35-38]</sup>等。

由于 Itti 的模型思路简单巧妙，且容易实现，效果也好，因而后来的研究者就在此基础上进行了相关的改进。在运算速度的提升方面，张鹏<sup>[39]</sup>等人提出了基于视点转移和视区追踪的图像区域检测算法，算法就是以此为基础的，而且速度上比它快了好多，接近一百倍。Matthew<sup>[40]</sup>实现了类似的思想，但采用的是人工智能中模拟智能结构的神经网络的方法，由于可以采用早期训练等原因，运算速度比 Itti 的方法更是快了 500 倍。Carota<sup>[41]</sup>用硬件实现了 Itti 的视觉注意计算模型，应用于机器人导航等机器视觉任务。在特征选择方面，Park<sup>[42,43]</sup>提出了专门为边缘和对称结构设计的新特征，其中有两个颜色特征图，对于识别边缘和对称结构有独特的效果。在注意单元的设计方面，Itti 提出的模型中是以竞争获胜的显著点为圆心、固定大小的长度为半径的圆作为视觉注意的单元的，这种固定大小的圆形区域没有考虑被注意物体的形状和大小，有时能够完全覆盖整个物体，但可能会因物体较小而同时覆盖其他无用的区域，有时则不能完全覆盖较大的物体，只能覆盖其中的一部分，因而会丢失相关信息。因此，为了解决这一问题，Satoh<sup>[44]</sup>、张鹏<sup>[29,45]</sup>分别提出了注意焦点的尺寸可以调整变化的视觉注意计算模型。在这些模型中，注意单元的圆形区域半径不是固定不变的，而是根据注意的目标物体的大小来决定的，注意焦点的大小由一定的算法预先计算得出，最终的注意焦点划住的区域和所要注意的目标物体的大小基本上差不多，但缺点是其形状仍为圆形，与真实的自然图像中的目标物体形状有很大的差别，注意区域不能根据目标物体形状的不同而改变。后来，Walther<sup>[46,47]</sup>通过分层反馈连接计算方法对模型进行了改进，使得注意区域可以实现随目标物体的大小和形状而自动改变，这样就使得模型计算得到的注意区域是一个实实在在的有意义的图像块。

通过以上对研究现状的分析，可以得出一个这样的结论：基于空间的视觉注意计算模型得到的结果是一个个的注意单元，这些所谓的显著区域不一定是有意义的，和实际的目标物体不一定对应，有时候还可能一点意义也没有。虽然有的模型能够实现注意区域的大小和形状随目标而改变，但由于先天的缺陷，是怎么也没有办法保证对应到真正完整的目标物体上的。

## 2) 基于物体的视觉注意计算模型

随着认知科学的发展，有越来越多的证据证明视觉注意是基于物体的注意<sup>[48,49]</sup>。在此基础上，研究者构建并发展出了不少基于物体的视觉注意计算模型<sup>[28,46,50-53]</sup>。在

基于物体的视觉注意计算模型中,最重要的一个概念就是“感知物体”,一般由格式塔(Gestalt)规则创建。但是鉴于格式塔规则是用心理学研究的术语描述的,不太好进行相应的计算,所以在基于物体的视觉注意计算模型中,“物体”的定义始终是一个最大的问题。

在对生物学实验进行大量观察后,Grossberg 和 Raizada<sup>[54]</sup>用神经网络模型构造了一个视觉编组模型,可以得到物体的轮廓。Sun 和 Fisher<sup>[28]</sup>首次提出的“分组”(grouping)概念,真正实现了基于物体的这种思想,对图像的位置、特征、结构物体、区域和物体组进行分层次分组操作,并结合了基于空间的视觉注意计算模型的一些基本原理,在模拟符合人类生理物理学实验的视觉注意和注意焦点的隐式转移方面有了重要突破。2008年,Sun<sup>[55]</sup>又扩展了这一模型,“分组”的概念进一步发展成为更具物理意义的“视觉物体”(visual-object)的概念,并且按照生物学成果对“视觉物体”进行自动且动态的知觉组织,实现了与人类视觉感知较为接近的对视觉物体的注意。

曾孝平等<sup>[56]</sup>提出基于图论的视觉注意模型,从图论的观点对目标物体表示进行了研究。Tian<sup>[57]</sup>提出的模型,试图使用感知颜色线索来进行基于物体的分割,只用颜色特征来分离物体,但在分离过程中没有考虑物体的大小、轮廓等其他因素,因而需要进行部件融合的工作,才能实现较为有意义的物体的注意。

### 3) 基于视觉通路的视觉注意计算模型

有关基于视觉通路的视觉注意计算模型的研究出现得较晚一些,最早的研究出现在1982年。通过研究猴子和人类大脑损伤后其视觉功能发生的变化,帮助人们认识到在视觉系统中,有两条不同的解剖学的信息通路,一条通路沿着大脑的腹部到达下颞叶皮层,另一条通路则沿背部到达后顶叶皮层。Ungerleider 和 Mishkin<sup>[58]</sup>首次给出这两条视觉通路的名称——“what”通路和“where”通路,并指出“what”通路负责处理“物体信息是什么”,即处理和表示物体特征,进行物体识别;“where”通路负责处理“物体信息在哪儿”,即处理和表示空间信息,进行空间定位。

比较经典的基于视觉通路的视觉注意计算模型是Rybak模型<sup>[59]</sup>,它包括3个子系统:低层子系统、中层子系统和高层子系统。其中,初级特征(如边缘信息的提取)在低层子系统中执行,类似于人类视觉系统中视网膜中央凹的工作;中层子系统负责生成不变性特征;“what”通路和“where”通路则在高层子系统中实现。田媚也构造了自己独特的模型<sup>[60]</sup>,首先根据环境信息提取一级“where”信息,然后通过Gabor滤波得到“what”信息和二级“where”信息,并形成自顶向下的注意控制,使模型有了一定的自顶向下的注意能力。Salah等<sup>[61]</sup>将可观测马尔可夫模型引入任务驱动的注意机制中,模型由3层组成。注意层模拟自底向上的初级注意;中间层是一个专家网络,由单层感知器组成,提取的信息类似“what”通路的信息;联合层则用离散的可观测马尔可夫模型连接下面两层的信息流,产生“where”通路的信息。

### 3. 视觉注意计算模型的研究现状分析（信息流角度）

此外，从视觉注意计算模型中的信息流角度分析，又可以分为自底向上和自顶向下两种研究方法<sup>[62]</sup>。

#### 1) 自底向上的视觉注意计算模型

自底向上的研究方法是从初级特征的提取开始，通过一定的数学变换，形成注意焦点，因而是一种数据驱动的工作方式。目前，绝大多数计算模型（尤其是前面提到的基于空间的视觉注意计算模型）都属于这类模型或是以此为基础而提出的改进模型，这里不再赘述。

#### 2) 自顶向下的视觉注意计算模型

自顶向下的研究方法强调高层知识对注意的指导作用，以任务为出发点，寻找有利于完成系统目标任务的特征，指导注意焦点的生成，因而是一种任务驱动的工作方式<sup>[63]</sup>。

肖洁<sup>[64]</sup>利用图斑引导感知分组过程，提出了一种对象积累的视觉注意计算模型，在一定程度上建立了高层语义和低层特征之间的联系。王慧<sup>[65]</sup>则综合空间和物体两个方面的信息，用一个控制机制去协同处理，实现两种机制的融合。综合这两个方面信息的模型还有窦燕提出的用于雕刻机器人的模型<sup>[66,67]</sup>。邵静<sup>[68,69]</sup>运用协同识别理论，研究了自底向上的信息和自顶向下的信息的融合机制，实现有目的的视觉注意。Rybak等<sup>[59]</sup>引入自顶向下的视觉控制参数进行物体检测和识别，且仅仅是作为线性组合与自底向上的注意机制相融合，离真正的自顶向下的注意机制仍有差距。Itti在后来的改进模型中，以心理阈值函数<sup>[26]</sup>的形式来控制视觉感知，在一定程度上实现了模拟自顶向下的视觉注意的效果。Peters和Itti<sup>[70-73]</sup>还利用眼动信息进行视觉注意点的预测，通过机器学习算法对被试者的眼动数据进行训练，得到主观的视觉注意点的位置模型，以自顶向下的方式引入预测过程中。侯晓迪和张丽清<sup>[74]</sup>提出了基于最大化视觉特征熵的动态视觉注意计算模型，用信息论的知识判断特征的显著性。

目前，自顶向下的研究方法虽然是注意模型研究中的热点，但研究成果相对较少，且思路仍过分依赖于自底向上的注意模型，因而没有实质性的突破。究其原因，是因为研究者对人类视觉系统如何生成自顶向下的注意能力的机制还不是很清楚。只有在对注意产生的机制甚至是智能产生的机制有更好的解释后，才可能带来质的飞跃。

## 4.2 基于特征加权的视觉注意计算模型

在视觉注意计算模型中,首先需要提取多个特征通道的信息,如颜色、朝向、亮度等特征,通过对其进行一系列的处理,得到各个特征维上的显著图,然后再将它们融合在一起,形成最终的兴趣图,在兴趣图中就可以按照一定的竞争机制,提取相应的显著区域,得到我们所关注的注意焦点。

一般情况下,各特征维上的显著图在融合过程中以相同的权重系数直接组合,或是通过提前指定特定的权重系数进行固定的线性组合。然而,对不同的物体而言,不同的特征对于其本身的重要性并不相同,就连同一物体的不同样本对于不同特征的依赖程度也不尽相同。

例如,一个颜色比较鲜艳的物体,其颜色特征相比朝向、亮度等特征而言就是最重要的;而一个棱角比较分明的物体,其朝向特征对其就显得尤其重要。而两个同样具有鲜艳颜色的不同物体,其不同样本中颜色的分布情况对于其也有十分重要的意义。如果一个物体的所有样本中的颜色呈现情况比较一致,则这种颜色就非常有助于表征该物体;而如果另一个物体的所有样本中的颜色呈现情况比较混乱,则颜色特征对其的表征作用也会受到一定程度的影响。

因此,如何根据物体样本自身的信息,动态生成其各个特征的权重系数,以便能更好地提取注意焦点,就变得十分有意义了。

通过对提取的不同特征通道上的显著图进行分析,本节设计了一种能够更好地表现不同特征通道对于目标物体的重要程度的特征加权算法,并据此对现有的经典视觉注意计算模型进行改进,提出了一种基于特征加权的视觉注意计算模型,该模型在相同条件下表现出了更好的效果<sup>[75]</sup>。

### 4.2.1 模型实现过程

#### 1. 提取底层特征

在提取底层特征方面,我们可以借鉴 Itti 视觉注意计算模型中的方法。

如前所述,Itti 视觉注意计算模型<sup>[26]</sup>是一种数据驱动的注意模型,主要基于 Treisman 特征整合理论。首先,从输入图像中提取多方面的特征,如颜色、朝向、亮度、运动等,运用一些数学工具对其进行处理,得到各个特征维上的显著图;然后,对这些显著图进行分析、融合得到兴趣图;最后,通过一定的竞争机制,从兴趣图中的多个待注意的候选区域中选出唯一的注意目标。

模型中采用线性滤波器组对图像滤波的方法,从图像中提取亮度、颜色、朝向等初级视觉特征。亮度和颜色特征用高斯滤波器进行采样,而朝向特征用 Gabor 滤波器得到。随后,将不同尺度的滤波器组合在一起,形成金字塔结构(也就是常说的高斯金字塔),就得到了同一幅图像的不同分辨率表示。

这样,就得到 1 个亮度特征图组、2 个颜色特征图组和 4 个朝向特征图组,共 7 组。每一组特征图组中有 9 层特征图。

对每一组特征图组执行中央周边差操作(记为 $\ominus$ ),就可以得到特征显著图。感受野的中央像素点位于金字塔的 $c$ 层,而周边像素点位于金字塔的 $s$ 层,其中 $s=c+\delta$ , $c \in \{2,3,4\}$ , $\delta \in \{3,4\}$ 。因此,对每一类特征,都可以得到 6 个特征显著图,分别为 2-5、2-6、3-6、3-7、4-7、4-8。最后,通过计算可以得到共 42 幅特征显著图,包括 6 幅亮度特征显著图、12 幅颜色特征显著图和 24 幅朝向特征显著图。

## 2. 计算特征权重

我们可以考察一下特征的分布:假如有一个物体,它有两个特征,其中某一特征的值分布比较密集,都集中到了一起,就说明这个特征对于识别这一物体是具有决定性作用的,或者说对于表征这一物体非常合适;而与之对应地,假设另外某一特征的值分布比较分散,且没有什么规律,就说明这个特征对于识别这一物体无足轻重,是不能体现物体特性的特征。

在视觉注意计算模型中,特征最后被量化成显著点的值。我们试想,如果一个物体在某个特征通道上得到的显著点的值变化很小,则说明这个特征对于识别这个物体很稳定,这样的特征对于识别这一物体非常有用,成为识别这一物体的首选判别依据。反之,如果某一特征的显著点的值变化范围很大,则说明此特征不能很好地表征这个物体,或者说此特征的描述方式不太恰当,这种特征对于识别该物体的作用不大,在识别这一物体时少用或不用它作为判别依据。因此,特征的重要性就可以用特征的分布情况来表征,具体地,可以用特征的标准差来表示。标准差是一个统计学的概念,通过计算集合中各个数据和它的平均值的距离的平均数来得到。标准差可以很好地表征数据分布的疏密程度,数据越密,标准差越小;相反,数据越疏,标准差越大。因此,用标准差可以非常方便地刻画特征分布情况,而且计算上也简单,容易实现,效果也非常好。

因为不同的特征通道计算出的显著值范围不同,要想在同一条件下比较不同的特征与物体的相关度,需要先对这些特征通道上计算出的显著值进行归一化,具体公式为

$$x'_l = \frac{x_l - m_l}{M_l - m_l} \quad (4-22)$$

其中, $x_l$ 表示某个物体第 $l$ 个特征通道(这里的特征通道指某一层次某一特征上的特征通道,如亮度特征的第二层通道等,总共有 42 个通道)上的特征值; $m_l$ 为该物体



第  $l$  个特征通道上显著值的最小值； $M_l$  为该物体第  $l$  个特征通道上显著值的最大值； $x'_l$  为  $x_l$  经过归一化之后的相应取值。

为了能够表征特征对于物体的重要程度，针对每一个特征关于物体的显著值分布，都可以用以下的计算方法具体计算。

假设计算出的显著值集中第  $j$  类共有  $n_j$  个样本，每个样本用  $L$  个特征来表示，其中第  $i$  个样本为  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iL}\}$ 。

设第  $j$  类第  $l$  个特征的标准差为  $\sigma_l$ ，即

$$\sigma_l = \sqrt{\sum_{i=1}^{n_j} (x_{il} - \bar{x}_l)^2 / (n_j - 1)} \quad (4-23)$$

其中， $x_{il}$  表示第  $j$  类中第  $i$  个样本的第  $l$  个特征； $\bar{x}_l$  则表示该类中所有样本第  $l$  个特征的平均值，定义为

$$\bar{x}_l = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{il} \quad (4-24)$$

这样就得到了该物体关于各个特征通道的分布。因为特征取值的分布越密集，标准差越小，表示该特征重要程度越高，可以看出，特征的重要程度与标准差之间成反比关系。

在这里，这种特征的重要程度可以定义为  $e_l$ ， $e_l \in [0, 1]$ ，即

$$e_l = \frac{1}{1 + \sigma_l} \quad (4-25)$$

从式 (4-25) 可以看出，标准差  $\sigma_l$  与重要程度  $e_l$  成反比关系，标准差  $\sigma_l$  越小，特征的重要程度  $e_l$  越大，表明该特征对于识别该物体越有用， $e_l$  也客观上成了计算不同特征的权重系数的依据。当图像的某个特征取值都集中地分布在某一点时，说明这个特征对于识别该目标物体至关重要，重要性极强，此时  $\sigma_l$  为 0， $e_l$  取到最大值 1。当  $\sigma_l$  取值较大时， $e_l$  取值较小，表明这个特征对于识别该目标物体几乎没有贡献，重要性较弱。

在进行具体图像特征通道权重计算时，由特征的重要程度得到特征权重系数（也称权值），第  $j$  类第  $l$  个特征的权重系数  $w_{jl}$  为

$$w_{jl} = \frac{e_l}{\sum_{l=1}^L e_l} \quad (4-26)$$

对于每一个物体，都可以得到一个权值矩阵，这个权值矩阵中包含了该物体的所有特征通道上的显著值的变化方式，也包含了这些特征对于识别该目标物体的贡献程度。把这个权值矩阵与由底层特征求得的底层特征（显著值）进行加权，即可得到更适合相应物体的特征值，这时的特征值已经不再是原始的显著信息，而是体现了由相应物体自身信息计算得来的关于不同特征对其重要性的权重信息，使得后边的特征显

著图的融合过程变得更加有针对性，效果也会更好。

### 3. 融合显著图

对于显著图的融合过程，可以借助 SalBayes 模型<sup>[76]</sup>（如图 4.7 所示）来说明。SalBayes 模型由 Elazary 等人提出，名字由 Saliency（显著性）和 Bayes（贝叶斯）两个单词合成而来，基于 Itti 视觉注意计算模型和贝叶斯理论，可以用来完成物体识别等任务。首先从输入图像中提取多方面的特征，如颜色、朝向、亮度等，形成各个特征维上的显著图。然后对同一物体不同样本在同一特征维度上的这些显著图分别进行分析，按照高斯分布进行特征维度上的独立建模，从而得到这一物体的一个表示（由不同特征维度上的一系列高斯组成）。当有新的测试样本时，用 Itti 视觉注意计算模型计算其各个特征维上的显著值，依据贝叶斯理论计算其相对于已经计算得到的物体表示的后验概率，后验概率最大者对应为目标物体，从而实现对物体的识别。

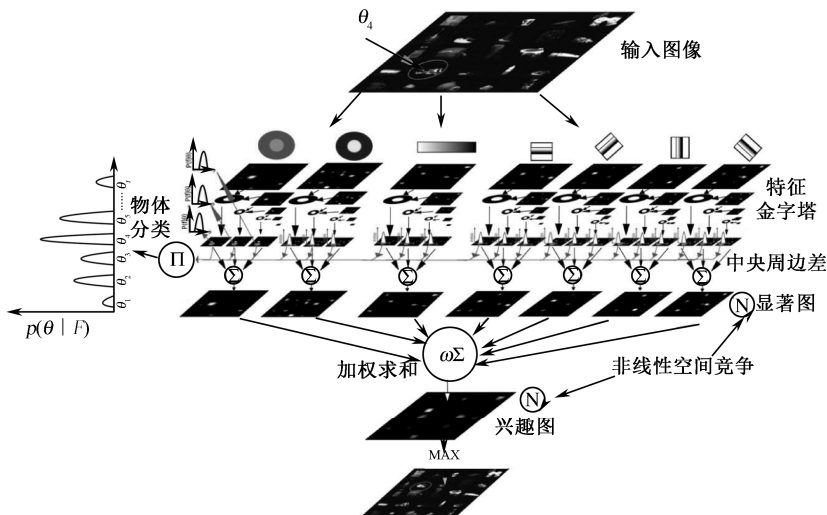


图 4.7 SalBayes 模型框图<sup>[76]</sup>

具体地，给定一个有  $N$  个像素点的兴趣区域块  $q$ ，对用前面的方法获得的 42 幅特征显著图中的任意一个，可以用一个空间竞争函数  $N(\cdot)$  来归一化，从而得到一个新的兴趣块  $q'$ ，我们取能使  $q'$  取得最大值时的  $q$  值来构造一个特征向量  $F$ 。特征向量  $F$  的第  $j$  个分量的值记为  $F_j$ 。

$$F_j = q \left[ \arg \max (q'_i)_{i=1,2,\dots,N} \right], \quad \forall j \in F \quad (4-27)$$

其中， $i$  代表兴趣块中的像素位置； $F$  是特征图集， $F_j$  是第  $j$  个特征图中的特征值。

把前面计算得到的特征权值  $w_{jl}$  加权到特征图集  $F$  上，得到新的特征图集，记为  $F^{\text{new}}$ ，这时的特征显著图已经不再是最初的特征显著图，而是融合了相应特征通道重要性的特征显著图。

$$F_j^{\text{new}} = F_j \cdot w_{jl} \quad (4-28)$$

接着, 用高斯分布对 ALOI 图像库的特征进行拟合, 其密度函数为

$$p(F_j^{\text{new}} | \theta_j) \propto N(F_j^{\text{new}}, \mu_j, \sigma_j) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left[ -\frac{(F_j^{\text{new}} - \mu_j)^2}{2\sigma_j} \right] \quad (4-29)$$

接下来, 使用朴素贝叶斯网络来对由特征显著图获得的特定特征进行分类, 具体公式为

$$p(\theta_j | F^{\text{new}}) = \frac{p(F_j^{\text{new}} | \theta_j) p(\theta_j)}{p(F^{\text{new}})} \quad (4-30)$$

要识别出物体, 需要对所有物体进行相应的计算, 然后计算其后验概率, 用最大的后验概率作为最终的结果。实验中的先验概率设置为  $1/C$ , 其中  $C$  是物体类别的数目, 取 1000。

由于概率密度可以作为概率归一常数, 所以可以把它从乘积中提出来, 即为

$$p(\theta_j | F^{\text{new}}) = p(\theta_j) \prod_{j=1}^n p(F_j^{\text{new}} | \theta_j) \quad (4-31)$$

此外, 在对大量概率值进行连乘操作时, 如果其中一些值非常小, 则很可能会导致最后的积发生溢出而不稳定。因此, 可以对概率进行取对数来解决这个问题。同时这也可以大大简化计算, 将原来的乘操作变为加操作, 而不会影响到最后的分类决策。通过这些不同的技术处理, 式 (4-30) 可以按式 (4-32) 计算。

$$p(\theta_j | F^{\text{new}}) = p(\theta_j) \sum_{j=1}^n \lg [p(F_j^{\text{new}} | \theta_j)] \quad (4-32)$$

根据式 (4-32), 可以对新来的测试样本计算亮度、颜色和朝向特征显著图, 并得到属于 ALOI 图像库<sup>[76]</sup>中每一个物体的概率。依据贝叶斯决策原理, 新来的测试样本属于概率最大的那个物体类别。

## 4.2.2 物体识别实验

### 1. 实验数据

为了验证提出的模型, 针对 ALOI 图像库<sup>[77]</sup>进行了物体识别任务测试。ALOI 图像库包含 1000 个物体的经过各种变化获得的照片。这些变化包括 12 种光照颜色、24 个光照角度和 72 个旋转角度, 对每一个物体而言, 都包含 108 张图片。图 4.8 是 ALOI 图像库中的一些示例图像。



图 4.8 ALOI 图像库中的一些示例图像

每个物体被 5 盏不同方向的灯照着，通过开关其中的一盏或两盏灯，以实现光照角度的变化，共形成 24 张光照角度不同的物体图片，如图 4.9 所示。

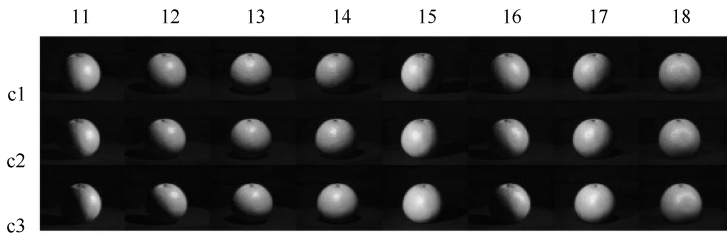


图 4.9 不同光照角度下采集到的系列图片

不同光照颜色下采集到的系列图片均为物体的正面，采集时开着 5 盏灯，光照颜色的温度变化范围为 2175K 到 3075K，照相机的白平衡值设为 3075K，共计 12 张图片，如图 4.10 所示。

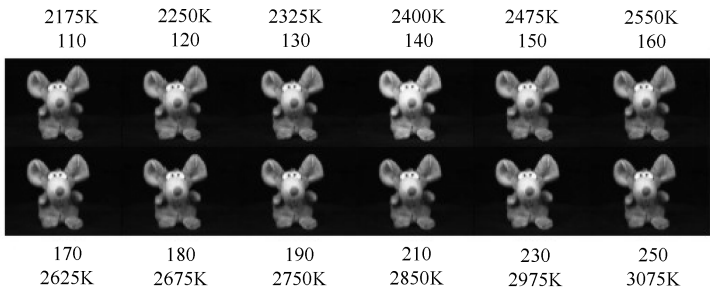


图 4.10 不同光照颜色下采集到的系列图片

对物体进行旋转，由一个固定摄像机进行拍摄，每旋转 5° 拍摄一张图片，形成因观察视角不同而体现物体特征的系列图片，旋转一周，共计 72 张图片，如图 4.11 所示。

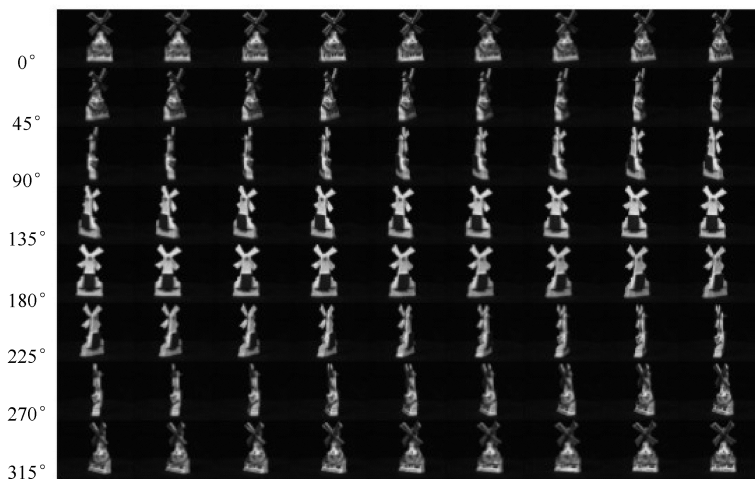


图 4.11 不同旋转角度下采集到的系列图片

## 2. 实验设置

实验中分别选取了不同数量的样本作为训练样本和测试样本，共设计了 4 个不同的实验数据集，分别记为数据集 1 至数据集 4，具体设置情况描述如下。

数据集 1：从 ALOI 数据库中每个物体的图像中分别随机选取 6.25% 的样本作为训练集，其余 93.75% 的样本作为测试集。

数据集 2：从 ALOI 数据库中每个物体的图像中分别随机选取 12.5% 的样本作为训练集，其余 87.5% 的样本作为测试集。

数据集 3：从 ALOI 数据库中每个物体的图像中分别随机选取 25% 的样本作为训练集，其余 75% 的样本作为测试集。

数据集 4：从 ALOI 数据库中每个物体的图像中分别随机选取 50% 的样本作为训练集，其余 50% 的样本作为测试集。

实验步骤如下。

S1：计算训练集中图像的特征显著图（42 维）。

S2：计算每一特征维的权重，并进行加权处理，得到新的特征显著图（42 维）。

S3：对所有物体在每一特征维上进行高斯拟合，得到所有物体的表示。

S4：计算测试集中图像的特征显著图（42 维）。

S5：依据贝叶斯理论计算其相对于已经计算得到的物体表示的后验概率。

S6：后验概率最大者对应为目标物体。

S7：统计正确识别的图像数目占测试集图像的百分比，即为模型的识别率。

S8：进行 5 次平行实验，取识别率的平均值作为最后结果。

SalBayes 模型作为基准实验，实验过程中不执行 S2 步骤。

### 3. 实验结果与分析

实验采用 Matlab R2010a 在 PC (CPU 为 Intel Pentium Dual Core 2.2GHz, 内存为 2GB) 上进行仿真, 代码参考自 Saliencytoolbox<sup>[78]</sup>, 并根据模型进行了相应修改。分别在 4 个数据集上用 SalBayes 模型<sup>[76]</sup>和新提出的特征加权模型进行对比实验, 在实验过程中随机选择样本, 分别进行 5 次平行实验, 最后求出平均值作为最终的识别率结果。比较两个模型的识别效果, 如图 4.12 所示。从图中可以看出, 在增加了根据物体自身信息得到的特征权重系数并用其对特征显著图融合过程进行调整后, 识别率有了一定提高, 体现出了新模型具有更好的注意能力。

由于 SalBayes 模型中 42 维特征的权重系数相同, 不同特征对物体显著性的重要性相同, 在融合过程中就会导致不重要的特征通道对重要的特征通道造成平滑影响, 即重要的特征通道的重要性被削弱甚至抵消, 这就使得用高斯拟合得到的物体表示的精度降低, 影响了识别率。而通过对特征通道的分布进行分析, 进而得出特征通道的重要性度量, 即特征权重系数, 并据此对特征通道进行调整后, 得到的物体表示就更加接近真实物体了, 因此特征加权模型的识别率较 SalBayes 模型会有所提高。

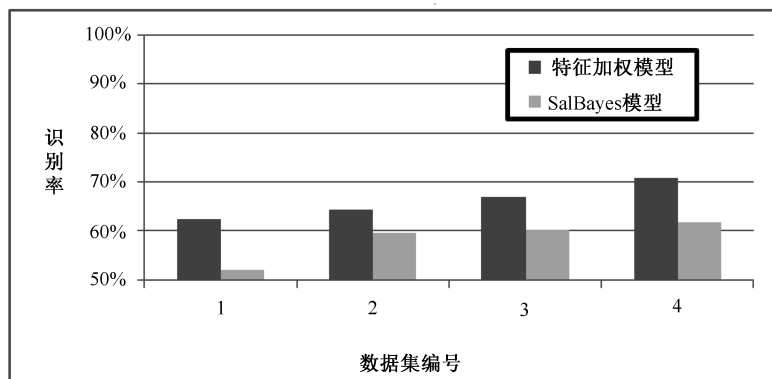


图 4.12 特征加权模型和 SalBayes 模型的识别率对比

#### 4.2.3 物体搜索实验

##### 1. 实验数据

该任务的实验数据库是用 ALOI 图像库生成的。从 ALOI 图像库中的 1000 个物体的 108 幅不同图片 (共计 108 000 幅) 中分别随机取 4 幅、9 幅、16 幅和 25 幅图片, 按  $2 \times 2$ 、 $3 \times 3$ 、 $4 \times 4$ 、 $5 \times 5$  排列, 各生成 1000 幅新的组合图像 (每一幅组合图像均由不同物体的图片拼接而成), 作为物体搜索任务数据库, 如图 4.13 所示。

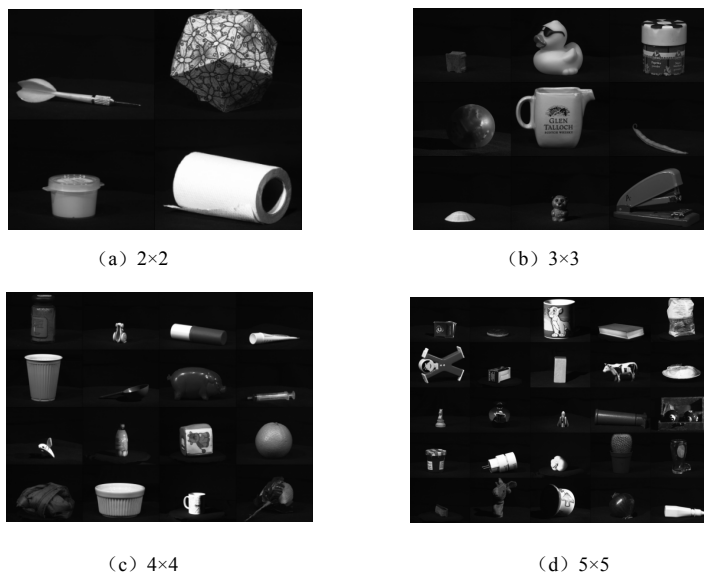


图 4.13 物体搜索任务数据库示例

## 2. 实验设置

实验任务是在组合图像中搜索指定的物体。实验步骤如下。

S1: 指定某一待搜索物体，用特征加权模型生成其表示。

S2: 计算含有此物体图像的组合图像的第一个显著点的特征向量。

S3: 用贝叶斯公式计算属于所有物体的概率，概率最大者为搜索结果。

S4: 利用注意转移机制，计算下一个显著点的特征向量。

S5: 对于 2×2 的组合图像集，重复执行  $n$  次（4 次）S3 和 S4，或执行到不再有新的显著点为止。

S6: 统计物体的正确识别率和正确识别物体时显著点的转移次数。

对于其他 3×3、4×4、5×5 的组合图像集，S5 的执行次数  $n$  分别取 9、16、25。

由于组合图像是随机选取 ALOI 库中的图像生成的，所以每一个 ALOI 库中的物体在组合图像库中出现的数目不完全相同，定义为  $N_i$ ，其中  $i=1,2,\dots,1000$ 。第  $j$  次正确识别出物体的次数为  $Y_j$ ，其中  $j=1,2,\dots,n$ ，则正确识别率  $P_i$  可由式（4-33）求得。

$$P_i = \frac{\sum_{j=1}^n Y_j}{N_i} \quad (4-33)$$

所有物体总的识别率为单个物体识别率的平均值，记为  $P$ ，计算公式为

$$P = \frac{\sum_{i=1}^C P_i}{C} \quad (4-34)$$

正确识别物体时显著点的转移次数记为  $t_{ij}$ 。其中,  $i=1,2,\dots,1000$ , 表示物体序号;  $j=1,2,\dots,n$ , 表示第  $j$  次正确识别出物体。正确识别某一物体时显著点的平均转移次数为  $t_i$ , 即

$$t_i = \frac{\sum_{j=1}^n t_{ij}}{N_i} \quad (4-35)$$

### 3. 实验结果与分析

图 4.14 所示为物体搜索任务的识别率。从图中可以看出, 组合尺寸为  $2 \times 2$  时, 正确识别率可以达到 70% 以上, 随着尺寸的增大, 识别率有了明显的下降, 原因是当组合图像中的物体太多时, 可能会导致因部分物体的显著性太低而不足以成为注意焦点的情况发生。

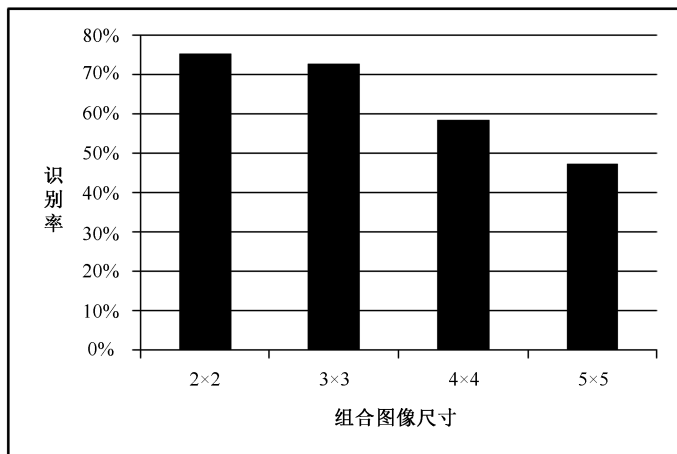


图 4.14 物体搜索任务的识别率



图 4.15 注意焦点无法转移到部分物体的示例

如图 4.15 所示, 注意焦点在转移 10 次后又回到了重复的显著点上, 导致其余物体没办法被成功找到。

图 4.16 所示为正确识别物体时显著点的转移次数统计图。图中给出了不同组合图像尺寸情况下, 显著点转移次数的统计。随着组合图像尺寸的增大, 转移次数的取值范围也被拉大, 也就是说, 在组合图像中的物体数目不断增加的情况下, 要搜索到相应物体, 需要进行的转移次数也随之增加。仔细分析可以看出, 在训练过程中,



将 ALOI 图像库中的各个物体进行单独训练，形成相应的物体表示，但在识别过程中，由于各个子图像形成的组合图像是一个整体，在计算其显著图时，各分子图像间会互相影响，因而使得正确识别物体时显著点的转移次数随着组合图像尺寸的增大而呈现增加的趋势。同时可以看出，转移次数增加的趋势会趋于平缓，说明各分子图像间的这种影响是有一定限度的，对于此类物体搜索任务的影响是可以接受的。

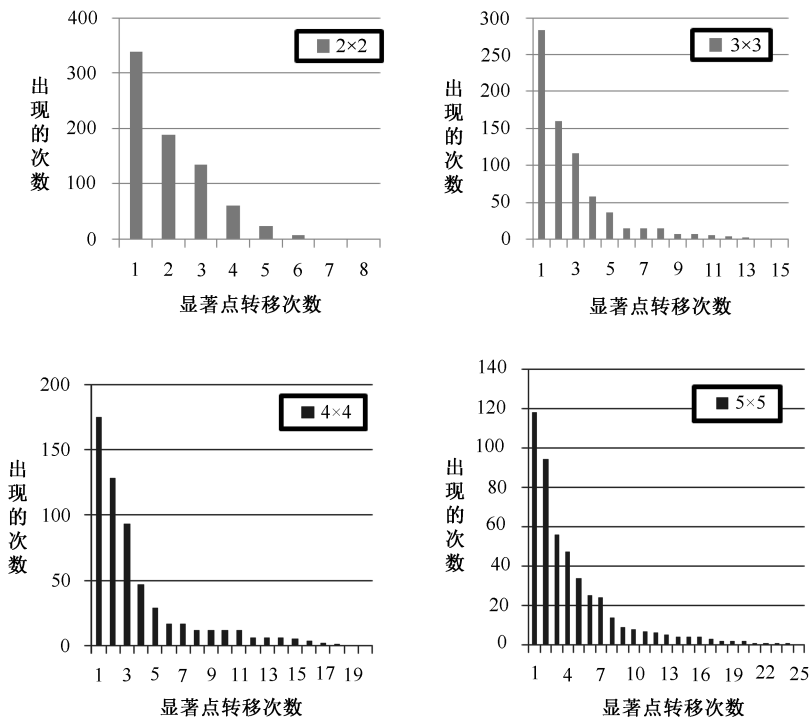


图 4.16 正确识别物体时显著点的转移次数统计图

### 4.3 基于高斯混合的视觉注意计算模型

人类在经过一定的训练之后，能够很容易地识别现实世界中的物体，而且在物体的形状、大小、观察角度、位置等发生变化的情况下，也能够轻松地完成识别任务。人们甚至能够一眼就认出一位老者年轻时候的照片，这说明人类视觉感知系统具有超强的鲁棒性。

生理学及心理学研究认为，人类感觉过程中会形成一个模板，当人们看到一个新的对象时，会生成一些神经元的刺激，在经过多次这种学习过程后，便形成了一个对应这个对象的表征模板，不同的对象会对应不同的表征模板，而在大脑中具体是如何

表征的，目前还没有很好的结论，这有待生理学家进行进一步的研究。

视觉注意计算模型可以将传统的物体识别过程和人类视觉加工机制相结合，形成一个合理的计算资源分配方案，引导整个物体识别处理过程，使待识别的物体能够快速有效地成为引起观察者注意的显著区域，使机器系统的物体识别处理过程具备类似人的视觉主动性和选择性。

然而对于同一个物体，人在不同角度下观察时，会得到不同的刺激，引起人注意的地方也不尽相同，但这些显著刺激仍能使人得到一个正确的识别结果，即轻松认出该物体。传统的视觉注意计算模型模拟了人类的视觉注意机制，但在针对同一物体的显著性建模方面没有进行深入的思考和研究。

本节研究了如何对视觉注意计算模型提取到的显著点进行分析，用什么样的统计方法来对物体进行建模，以及建什么样的模型合适的问题。针对物体识别任务，本节讨论了现有视觉注意计算模型中所用单高斯模型的不足，从理论上分析了使用混合高斯模型（也称高斯混合模型）的充分性和必要性，并在 ALOI 数据集<sup>[77]</sup>上进行了实验验证，从识别率等方面对所提出的模型与原有模型进行对比分析，结果表明效果得到了有效改善。

### 4.3.1 高斯混合模型

我们知道，不管是机器还是人，学习的过程都可以看作一种“归纳”的过程，在归纳的时候，需要有一些假设的前提条件。例如，当被告知水里游的那个家伙是鱼之后，使用“在同样的地方生活的是同一种东西”的假设，可以归纳出“在水里游的都是鱼”这样一个结论。当然这个过程是完全“本能”的，如果不仔细去想，我们也许不会了解自己是怎样“认识鱼”的。另一个值得注意的地方是这样的假设并不总是完全正确的，甚至可以说总是会有这样那样的缺陷，因此我们有可能把虾、龟，甚至是潜水员当作鱼。也许我们觉得可以通过修改前提假设来解决这个问题，例如，基于“生活在同样的地方并且穿着同样衣服的是同一种东西”这个假设，得出结论：在水里游并且身上长有鳞片的是鱼。可是这样还是有问题，因为那些没有长鳞片的鱼现在又被排除在外了。

在机器学习研究中，对于数据的假设有很多种，根据不同的背景和知识，选择不同的模型，从而使得学习效果最接近于真实的数据集，达到最好的学习效果。应用比较广泛的是高斯分布假设，这是一种最实用的分布假设，世界上的大部分数据都近似服从这一分布，可以在对数据没有其他认识的情况下很好地模拟数据集的分布。

高斯分布（gaussian distribution）是一个在数学、物理及工程等领域都非常重要的连续概率分布函数，它描述了一种围绕某个单值聚集分布的随机变量。在生活中，各种各样的心理学测试分数和物理现象，如人的身高、寿命等，都被发现近似地服从高

斯分布。同时，高斯分布也是统计学及许多统计测试中最广泛应用的一类分布。由于高斯分布的广泛性和熵最值性等优点，所以很多模型中都采用这一最简单也最有效的分布作为其分布假设。但是在有些情况下，当所要表征的数据并不服从高斯分布假设时，就不能很好地描述对象，甚至会造成错误。为此，有研究者根据对一些对象数据的分析后，提出了混合高斯分布、超高斯分布等其他复杂的分布假设，以便能尽可能准确地表征所要描述的对象。

虽然我们可以用不同的分布来随意地构造混合模型，但是高斯混合模型（Gaussian Mixture Model, GMM）<sup>[79]</sup>最为流行。顾名思义，高斯混合分布就是数据可以看作从多个高斯分布中生成的，如图 4.17 所示。

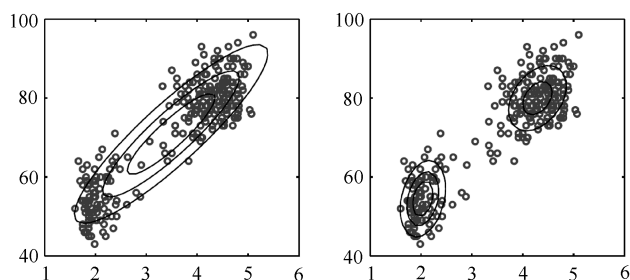


图 4.17 高斯混合分布示意图

另外，混合模型本身其实也是可以变得任意复杂的，通过增加参加混合的模型的个数，我们可以任意地逼近任何连续的概率密度分布。

每个 GMM 由  $K$  个高斯分布组成，每个高斯分布称为一个“组件”，这些组件线性加成在一起就组成了 GMM 的概率密度函数，即

$$p(x) = \sum_{k=1}^K p(k)p(x|k) = \sum_{k=1}^K \pi_k N(x|\mu_k, \sigma_k) \quad (4-36)$$

根据式 (4-36)，如果我们要从 GMM 的分布中随机地取一个点，实际上可以分为两步：首先，随机地在这  $K$  个组件之中选一个，每个组件被选中的概率实际上就是它的系数  $\pi_k$ ；其次，在选中了组件之后，再单独地考虑从这个组件的分布中选取一个点就可以了——这里已经回到了普通的高斯分布，转化为了已知的问题。假设现在有  $N$  个数据点，我们认为这些数据点由某个 GMM 产生，现在需要确定  $\pi_k$ 、 $\mu_k$ 、 $\sigma_k$  这些参数。自然地，我们想到利用最大似然估计来确定这些参数，GMM 的似然函数为

$$\lg \prod_{i=1}^N p(x_i) = \sum_{i=1}^N \lg p(x_i) = \sum_{i=1}^N \lg \sum_{k=1}^K \pi_k N(x_i|\mu_k, \sigma_k) \quad (4-37)$$

由于在对数函数里面又有加和，所以我们没法用求导解方程的办法直接求得最大值。为了解决这个问题，我们利用期望-最大化（Expectation Maximization, EM）算法，分步迭代，进而求得最大值，并求出取得最大值时各个参数的值。具体计算过程可以分为以下几个步骤。

(1) 初始化参数  $\pi_k$ 、 $\mu_k$ 、 $\sigma_k$  的一种流行的做法是先通过  $k$  均值算法对数据点进行聚类, 根据聚类结果选取参数的初始值。

(2) E 步: 求期望。估计数据由每个组件生成的概率 (并不是每个组件被选中的概率)。对每个数据  $x_i$  来说, 它由第  $K$  个组件生成的概率为

$$\gamma(i, k) = \frac{\pi_k N(x_i | \mu_k, \sigma_k)}{\sum_{j=1}^K \pi_j N(x_i | \mu_j, \sigma_j)} \quad (4-38)$$

然而, 由于式 (4-38) 里的  $\pi_k$ 、 $\mu_k$ 、 $\sigma_k$  正是需要我们估计的参数, 因此采用迭代法, 即取上一次迭代所得的值 (或者初始值)。

(3) M 步: 最大化。对式 (4-38) 进行求导, 求出最大似然所对应的参数值

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^N \gamma(i, k) x_i \quad (4-39)$$

$$\sigma_k = \frac{1}{N_k} \sum_{i=1}^N \gamma(i, k) (x_i - \mu_k)(x_i - \mu_k)^T \quad (4-40)$$

其中,  $N_k = \sum_{i=1}^N \gamma(i, k)$ 。根据条件可知, 参数  $\pi_k$  满足  $\sum_{i=1}^N \pi_k = 1$ , 因此我们在 GMM 的似然函数中加入拉格朗日乘子  $\lg \sum_{i=1}^N p(i, k) + \lambda (\sum_{i=1}^K \pi_k - 1)$ , 求得该式取得最大值时  $\pi_k$  对应的值, 即

$$\pi_k = \frac{N_k}{N} \quad (4-41)$$

(4) 计算似然函数的值[由式 (4-37) 可得], 检查似然函数是否收敛。若收敛了, 说明似然函数已经取得最大值, 此时参数对应的值即为各参数的最大似然估计。否则, 继续进行 E 步和 M 步的迭代。

### 4.3.2 基于 GMM 的视觉注意计算模型

#### 1. 高斯混合的充分性和必要性

##### 1) 充分性

如前所述, 混合模型可以通过调整参与混合的模型的类型、个数及混合比例等一些参数, 任意地逼近任何连续的概率密度分布。因此, 理论上, 用混合模型可以更好地实现对数据的建模, 使得所用模型的物体识别效果更好。当然这样一来, 会使得模型变得更加复杂, 计算量会相应增加, 因此, 实际建模过程中, 需要根据实际问题进行折中处理, 既要能够对物体进行合适的表征, 又要使模型在可控的时空范围内有效。

在可调整的若干参数中,最重要的是参与混合的模型的类型。不同的概率分布可以对不同数据进行估计,效果取决于所使用的概率分布是否与真实的数据分布相吻合,如果所使用的概率分布与实际数据分布大致接近,则会得到好的识别结果,反之,则得不到好的结果,还可能会得到错误的结果。因此,选择混合模型的类型时,需要对数据进行充分的分析。一般情况下,数据的分布基本服从高斯分布,且高斯分布在计算上具有比其他分布更多的优势。因此,在没有其他额外信息的情况下,将混合模型的参与者设计成简单高效的高斯分布是可行的、充分的。

## 2) 必要性

现实世界中的物体以各种不同的姿态呈现在我们的视觉系统中,形成了一个相对稳定的物体模板,其实际样本的多样性使得用单一的高斯假设变得不合理,成为识别效果不好的一个重要原因。而传统的视觉注意计算模型对图像数据中待识别物体的建模就是采用了最简单的高斯分布假设或者干脆不作假设,仅凭视觉显著点进行物体的检测和识别,没有充分利用待识别物体在现实世界中本身的信息,使得现有的视觉注意计算模型的识别率不高,且没有真正体现视觉注意机制的智能性。

对图像数据中物体的建模,因其数据的高维性而变得比较复杂,再加上大千世界中的物体繁多,想要设计一种统一的数学模型对其表征是不现实的。因此,寻求一种既能较好地表征现实物体又具有现实可行性的物体表征方法,就显得尤为重要。

现实物体的非高斯性源于其多面性,所谓“横看成岭侧成峰”,人在不同的角度观察物体会有不同的感受,我们之所以能认为其为同一物,源于我们事先分别用“岭”和“峰”对其进行了建模,然后综合这两者,得到了一个统一的“庐山真面目”。

例如,ALOI 图像库中每个物体的图片都是经过变换各种精心设计好的光照颜色、光照角度或旋转角度而获得的。而这些变化使得对应于每个物体的一系列图片被系统地分成了几个部分,体现为数据上强烈的非高斯性。为了能更好地表征现实世界中的物体,在对其进行建模时,使用混合高斯模型就变得十分必要了。

## 2. 用视觉注意机制提取视觉显著图

本小节所用到的视觉注意计算模型源于 Elazary 等人提出的 SalBayes 模型<sup>[76]</sup>。这是一种数据驱动的注意模型,主要基于 Itti 视觉注意计算模型和贝叶斯理论<sup>[79]</sup>。首先从输入图像中提取多方面的特征,如颜色、朝向、亮度等,形成各个特征维上的显著图。然后对同一物体的这些显著图分别进行分析,按照高斯分布进行建模,从而得到这一物体的一个表示。当有新的测试样本时,用 Itti 视觉注意计算模型计算其各个特征维上的显著值,依据贝叶斯理论计算其相对于已经计算得到的物体表示的后验概率,实现对物体的识别。

提取亮度、颜色、朝向等初级视觉特征的过程完全和 Itti 视觉注意计算模型相同。亮度和颜色特征提取采用高斯滤波器,而朝向特征采用 Gabor 滤波器。低通滤波采用

的高斯模板大小为  $5 \times 5$ 。用于实现非均匀采样的高斯金字塔结构共分 9 层，不同层间相减模拟了中央周边差采样方式。

最后，通过计算可以得到共 42 幅特征显著图，包括 6 幅亮度特征显著图、12 幅颜色特征显著图和 24 幅朝向特征显著图。

### 3. 用高斯混合模型对物体显著点进行建模

对同一物体的不同图像分别计算各个特征显著图上的显著点，对其进行高斯混合建模。物体建模分为训练过程和测试过程两个阶段。具体模型框图如图 4.18 所示。

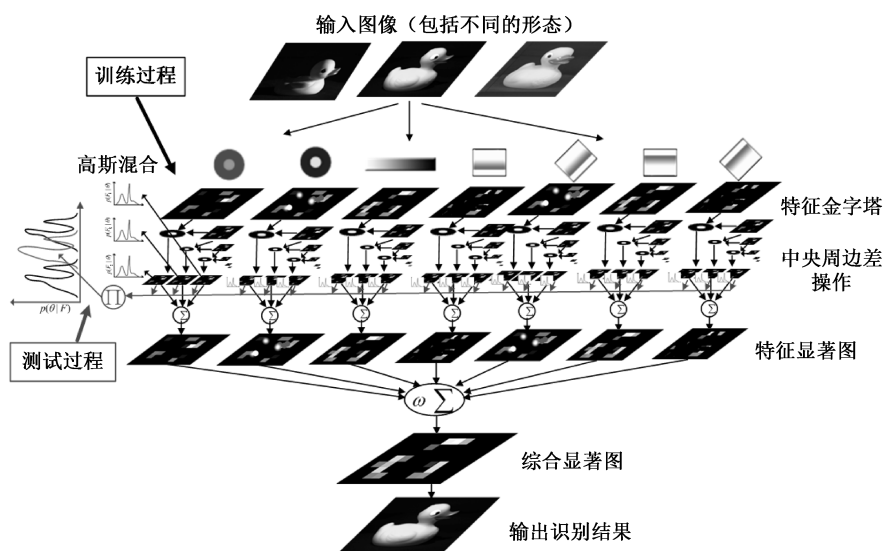


图 4.18 基于 GMM 的视觉注意计算模型框图

#### 1) 训练过程

在训练过程中，通过计算 42 幅特征显著图的概率分布来建立对物体表征模板的描述。其中，概率密度函数使用 3 个高斯分布的均值和方差及相应的混合比例来设置。也就是说，该模型针对 42 幅特征显著图，分别为每个物体学习 3 个单独的高斯分布，然后再将它们混合在一起。

用类似前面的方法，对用上面提到的方法获得的 42 幅特征显著图中的任意一个，构造一个特征向量  $\mathbf{F}$ 。特征向量  $\mathbf{F}$  的第  $j$  个分量的值记为  $F_j$ 。

$$F_j = q \left[ \arg \max (q'_i)_{i=1,2,\dots,N} \right], \quad \forall j \in F \quad (4-42)$$

其中， $i$  代表兴趣块中的像素位置； $F$  是特征图集； $F_j$  是第  $j$  个特征图中的特定的特征值； $q$  为含有  $N$  个像素点的兴趣区域块； $q'$  是用空间竞争函数  $N(\cdot)$  归一化后得到的新兴趣块。

由于 ALOI 图像库系统地分为 3 个部分,所以可以用  $k$  均值方法来对  $F_j$  进行聚类,类别分别为  $k$ , 其中  $k \in \{1, 2, 3\}$ 。当给定一个物体特征  $\theta_j$  时,对其中的每一个类别  $F_{jk}$  均用高斯分布来建模,其密度函数为  $p(F_{jk} | \theta_j)$ , 即

$$p(F_{jk} | \theta_j) \propto N(F_{jk}, \mu_{jk}, \sigma_{jk}) = \frac{1}{\sigma_{jk} \sqrt{2\pi}} \exp \left[ -\frac{(F_{jk} - \mu_{jk})^2}{2\sigma_{jk}} \right] \quad (4-43)$$

最终的物体表征模板  $\theta$  可以用针对每一个单独特征显著图的参数  $\theta_j$  的集合来表示。每一个参数由 3 个均值  $\mu_{jk}$ 、3 个标准差  $\sigma_{jk}$  和 3 个混合比例  $\pi_{jk}$  构成。

$$p(F_j | \theta_j) = \sum_{k=1}^3 \pi_{jk} p(F_{jk} | \theta_j) \quad (4-44)$$

由此,得到了对物体的表示。

## 2) 测试过程

测试过程使用朴素贝叶斯网络来对由特征显著图获得的特定特征进行分类,具体公式为

$$p(\theta_j | F_j) = \frac{p(F_{jk} | \theta_j) p(\theta_j)}{p(F_j)} \quad (4-45)$$

要识别出物体,需要对所有物体进行相应的计算,然后计算其后验概率,用最大的后验概率作为最终的结果。实验中的先验概率设置为  $1/C$ , 其中  $C$  是物体类别的数目,取 1000。

由于概率密度可以作为概率归一常数,所以可以把它从乘积中提出来,即为

$$p(\theta_j | F_j) = p(\theta_j) \prod_{j=1}^n p(F_{jk} | \theta_j) \quad (4-46)$$

此外,在对大量概率值进行连乘操作时,如果其中的一些值非常小,则很可能会导致最后的积发生溢出而不稳定。因此,可以对概率进行取对数来解决这个问题。同时这也可以大大简化计算,将原来的乘操作变为加操作,而不会影响到最后的分类决策。通过这些不同的技术处理,式(4-45)可以按式(4-47)计算。

$$p(\theta_j | F_j) = p(\theta_j) \sum_{j=1}^n \lg [p(F_{jk} | \theta_j)] \quad (4-47)$$

根据式(4-47),可以对新来的测试样本计算亮度、颜色和朝向特征显著图,并得到属于 ALOI 图像库中每一个物体的概率。依据贝叶斯决策原理,新来的测试样本属于概率最大的那个物体类别。

### 4.3.3 实验与分析

#### 1. 实验一结果及分析

本实验针对 ALOI 图像库中的 3 类变换方式——光照颜色、光照角度和旋转角度，将 ALOI 图像库分为 3 个子集，然后用 SalBayes 模型分别进行测试，以测试在不同子集上模型的识别率，其识别率如图 4.19 所示。

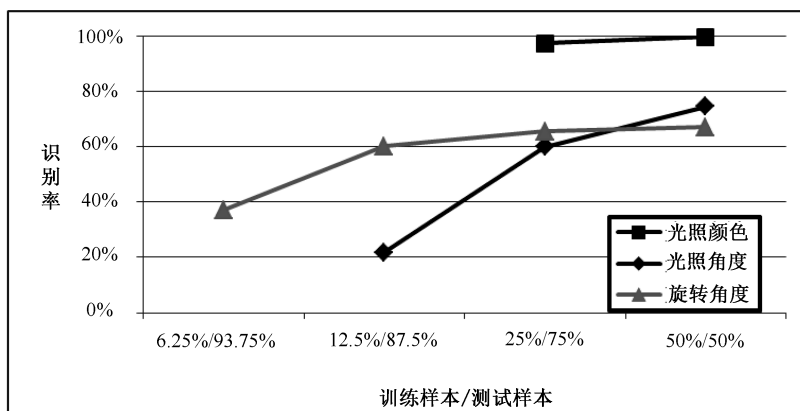


图 4.19 针对不同变换方式的识别率

由于 ALOI 图像库的 3 个子集的图像数量不同（分别为：光照颜色子集为 12、光照角度子集为 24 和旋转角度子集为 72），因此实验的训练集和测试集的设置也有所不同。

对光照颜色子集而言，设置两种实验数据集，分别是：从 ALOI 数据库的光照颜色子集中每个物体的 12 幅图像中分别随机选取 25% 的样本作为训练集，其余 75% 的样本作为测试集；另一种设置是取 50% 的样本作为训练集，其余 50% 的样本作为测试集。

对光照角度子集而言，设置 3 种实验数据集，分别是：从 ALOI 数据库的光照角度子集中每个物体的 24 幅图像中分别随机选取 12.5% 的样本作为训练集，其余 87.5% 的样本作为测试集；另两种设置分别是取 25%、50% 的样本作为训练集，其余 75%、50% 的样本作为测试集。

对旋转角度子集而言，设置 4 种实验数据集，分别是：从 ALOI 数据库的旋转角度子集中每个物体的 72 幅图像中分别随机选取 6.25% 的样本作为训练集，其余 93.75% 的样本作为测试集；另 3 种设置分别是取 12.5%、25%、50% 的样本作为训练集，其余 87.5%、75%、50% 的样本作为测试集。

实验参考自 Saliencytoolbox<sup>[78]</sup>，并根据实验设置进行了相应修改。



从图 4.19 中可以看出, SalBayes 模型对光照颜色不敏感, 对由这种变化产生的不同图像都能很好地识别, 识别率很高, 从实验数据看, 即使只使用 3 个样本作为训练集, 训练出的模型也能达到 97.2% 的正确识别率。但是由光照角度和旋转角度不同造成的样本不同, 会使得物体的图像产生比较大的视觉上的变化, 同样只使用四分之一样本作为训练集时, 识别率便直线下降, 只有 60% 多, 即使是将一半样本作为训练集, 识别效果也不是特别理想。图中的数据显示模型对旋转角度更加敏感, 原因是旋转后的图像的差别更大。

## 2. 实验二结果及分析

本实验针对 ALOI 全数据集, 分别用 SalBayes 模型和 GMM 改进后的模型进行实验对比, 比较两个模型的识别性能。

实验中分别选取了不同数量的样本作为训练样本和测试样本, 共设计了 4 个不同的实验数据集, 分别记为数据集 1 至数据集 4, 具体设置情况描述如下。

数据集 1: 从 ALOI 数据库中每个物体的图像中分别随机选取 6.25% 的样本作为训练集, 其余 93.75% 的样本作为测试集。

数据集 2: 从 ALOI 数据库中每个物体的图像中分别随机选取 12.5% 的样本作为训练集, 其余 87.5% 的样本作为测试集。

数据集 3: 从 ALOI 数据库中每个物体的图像中分别随机选取 25% 的样本作为训练集, 其余 75% 的样本作为测试集。

数据集 4: 从 ALOI 数据库中每个物体的图像中分别随机选取 50% 的样本作为训练集, 其余 50% 的样本作为测试集。

实验步骤如下。

S1: 计算训练集中图像的特征显著图 (42 维)。

S2: 对所有物体在每一特征维上进行混合高斯拟合, 得到所有物体的表示。

S3: 计算测试集中图像的特征显著图 (42 维)。

S4: 依据贝叶斯理论计算其相对于已经计算得到的物体表示的后验概率。

S5: 后验概率最大者对应为目标物体。

S6: 统计正确识别的图像数目占测试集图像的百分比, 即为模型的识别率。

S7: 进行 5 次平行实验, 取识别率的平均值作为最后结果。

SalBayes 模型作为基准实验, 实验过程中 S2 步骤中的混合高斯拟合改为单高斯拟合。

实验结果如图 4.20 所示。从图中可以看出, 在 4 个数据集上, 基于高斯混合的模型比原模型的识别率都有所提高。这是因为每个物体的不同样本间存在一定的差别, 用单高斯拟合时, 没有充分体现实际的样本构成, 而用高斯混合进行拟合后, 缩小了

拟合模型与实际样本间的误差，从而提高了识别率。当然，按照前面的分析，还可以设计更复杂的混合模型，但是相应的计算复杂度也会增加，而且对不同的数据集而言，样本分布也不相同，因此这里需要根据实际情况进行合理设计，在计算复杂度和识别率之间折中处理，才能得到最佳的效果。

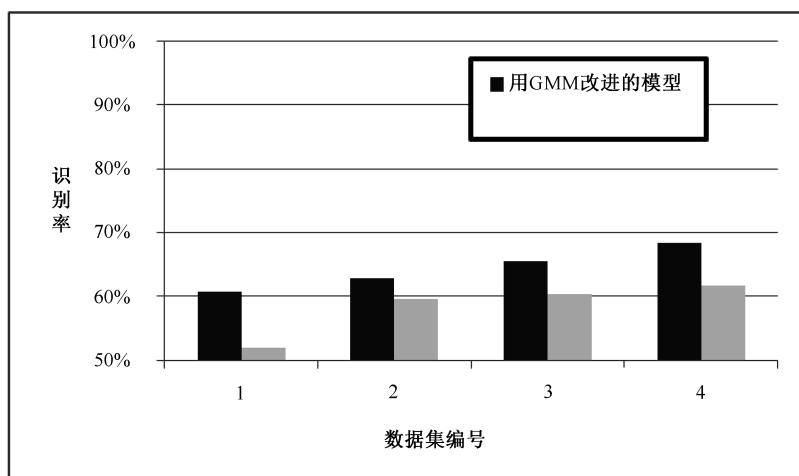


图 4.20 原模型和改进后的模型的对比

通过对识别率前十名和后十名的物体（如图 4.21 所示）进行分析，发现前十名的物体具有轴对称性，旋转角度对其图像的影响较小，而相反，后十名的物体旋转后视觉差异较大，这也同时验证了实验一中分析得到的结论。

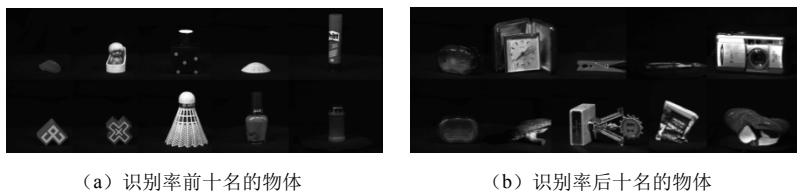


图 4.21 识别率前十名和后十名的物体

图 4.22 是识别率第一名的物体（贝壳）的亮度显著值分布。从图中可以看出，108 幅图片的显著值差别不大，其均匀的颜色和良好的轴对称性使其在光照颜色、光照角度及旋转角度等变化下仍保持着较好的视觉一致性，从而取得了最高的识别率。

图 4.23 是识别率后十名中的某物体（收音机）的亮度显著值分布。从图中可以看出，由于收音机表面有光滑的部分，也有比较粗糙的部分，且整体呈现长方体，外形极不规则，因此导致 108 幅图片的显著值差别非常大，视觉一致性较差，识别率很低。

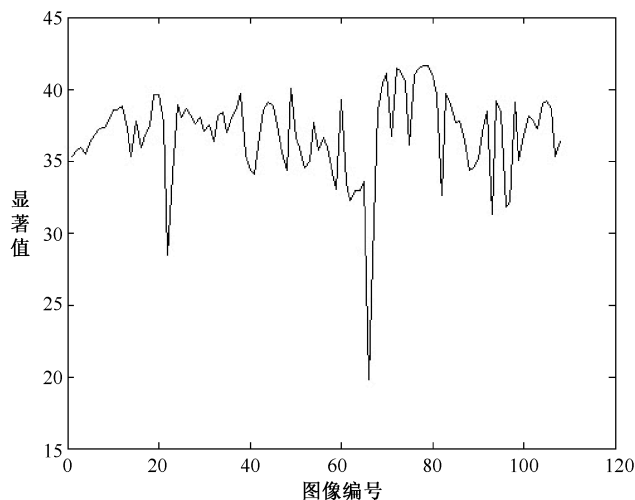


图 4.22 识别率第一名的物体的亮度显著值分布

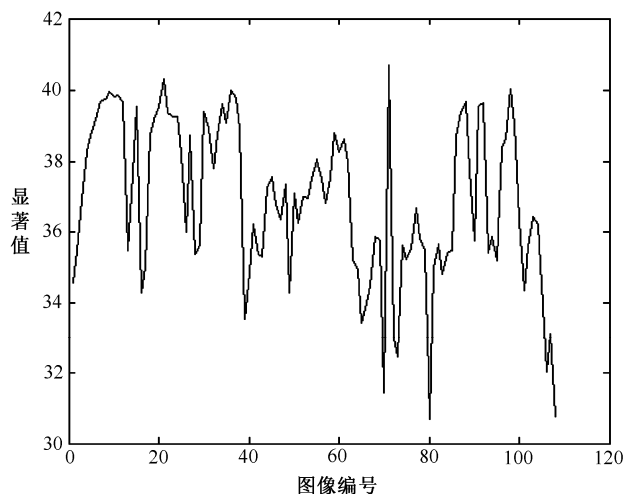


图 4.23 识别率后十名中某物体的亮度显著值分布

## 4.4 基于 CIELab 的视觉注意计算模型

在人类视觉系统接收到的众多刺激中,颜色是最重要的一类。颜色是通过眼、脑和我们的生活经验所产生的一种对光的视觉效应<sup>[80]</sup>。众所周知,颜色对人的心理和生理影响很大,因此,在研究视觉注意计算模型时,对颜色特征需要给予更多的关注。

颜色模型(或颜色空间)就是在一定的标准下,用一定的数学形式对颜色加以说明,关于颜色空间的描述,详见第2章。其中 CIELab 颜色模型是用来描述人眼可见的所有颜色的最完备的色彩模型之一,在设计思路与人类视觉系统对颜色的感知机

制最为吻合<sup>[81]</sup>。但遗憾的是,目前的视觉注意计算模型多以 RGB 颜色模型作为特征。因此,本节提出了一种基于 CIELab 的视觉注意计算模型,将原模型中的颜色特征空间由 RGB 空间转换到 CIELab 空间,并对特征进行了一些独到的设计,使模型对颜色对比信息更为敏感,并在禁止返回机制中加入了自适应阈值控制,使模型对样本的颜色变化能动态适应,还在模型的显著区域识别阶段,采用尺度不变性特征转换 (Scale-Invariant Feature Transform, SIFT) 特征<sup>[82]</sup>确定其有效性。该模型在针对拥有丰富色彩的物体(如交通标志)的检测与识别任务中取得了很好的效果。

#### 4.4.1 模型实现过程

Itti 等人提出的视觉注意计算模型,以特征整合理论为基础,采用了自底向上和基于空间的控制策略,而且通过仿生物视觉系统的神经机制来实现。该模型以亮度、颜色和朝向作为引导视觉注意的早期视觉特征,其计算的核心有两点:一是通过建立高斯金字塔来模仿中央周边差操作,这也是视网膜、外膝体及视皮层内采用的一般计算原则;二是通过使用高斯差模型来模拟感受野机制,得到各类特征的显著图,并由各类特征显著图之间的线性组合形成兴趣图,最后根据兴趣图得到注意焦点。模型中的颜色特征选用 RGB 空间,由于目前大部分图像格式或视频采集设备都采用 RGB 格式,因此体现出了其使用方便的优点。但正如前文所述,RGB 空间模型是面向硬件的彩色模型,因而最适用于彩色监视器的显示或视觉信息捕捉设备的成像,并不真正符合人类视觉系统对于颜色的感知机制。因此,本节提出的模型对原模型进行了调整,在特征选择上保留了颜色和亮度通道,将特征空间从 RGB 空间转换到 CIELab 空间,并在禁止返回机制中加入了自适应阈值控制,使算法能避免在非显著区域停留,并消除了环境光线对物体检测过程的影响。模型的识别部分主要采用模板匹配算法,首先提取标准图像库的 SIFT 特征作为模板,然后对由视觉注意计算模型得到的待识别图像显著区域进行同样的 SIFT 特征提取,并计算与模板的相似度,找出相似度最大的作为识别结果。基于 CIELab 的视觉注意计算模型的流程图如图 4.24 所示。

##### 1. 提取基于 CIELab 的颜色特征

CIELab 颜色空间是颜色对立空间,设计思路接近人类视觉感知。它分别有 3 个维度,其中  $L^*$  表示亮度 ( $L^*=0$  为黑色而  $L^*=100$  为白色),  $a^*$  和  $b^*$  表示颜色对立维度 ( $a^*=-50$  为绿色而  $a^*=+50$  为品红,  $b^*=-50$  为蓝色而  $b^*=+50$  为黄色)。CIELab 颜色空间致力于感知均匀性,其中  $L^*$  分量密切匹配人类亮度感知,因此可以通过修改  $a^*$  和  $b^*$  分量的输出色阶来作精确的颜色平衡,或使用  $L^*$  分量来调整亮度对比。这些变换在 RGB 等其他颜色空间中是困难或不可能的,因为 RGB 等是用来建模物理设备的输出,而不是人类视觉感知<sup>[83]</sup>。

具体地,在视觉注意计算模型中,先将输入的图像由 RGB 模型转换到 CIELab 模

型, 分别用  $r$ 、 $g$ 、 $b$  来表示输入图像的 3 个颜色分量,  $L^*$ 、 $a^*$ 、 $b^*$  分别为转换之后的 3 个分量, 具体转换方法见第 2 章。

在此基础上, 建立 4 个宽调谐的颜色通道, 为了区分基于 RGB 的视觉注意计算模型中的颜色通道, 这里分别记为红色  $R^*$ 、绿色  $G^*$ 、蓝色  $B^*$  和黄色  $Y^*$ , 计算公式分别为式 (4-48) 至式 (4-51)。

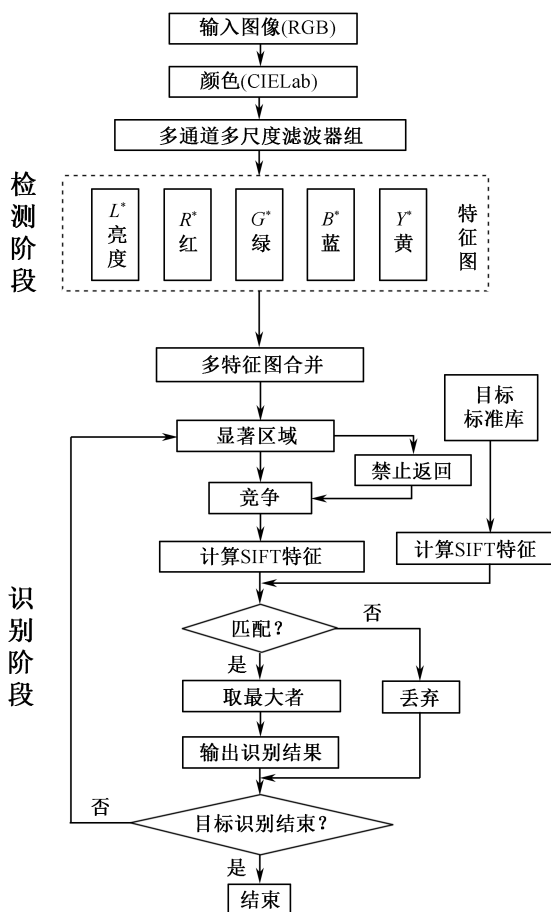


图 4.24 基于 CIELab 的视觉注意计算模型的流程图

$$R^* = \begin{cases} \frac{a^*}{L^*}, & a^* > t_a \\ 0, & a^* \leq t_a \end{cases} \quad (4-48)$$

$$G^* = \begin{cases} -\frac{a^*}{L^*}, & a^* > t_a \\ 0, & a^* \leq t_a \end{cases} \quad (4-49)$$

$$Y^* = \begin{cases} \frac{b^*}{L^*}, & b^* > t_b \\ 0, & b^* \leq t_b \end{cases} \quad (4-50)$$

$$B^* = \begin{cases} -\frac{b^*}{L^*}, & b^* > t_b \\ 0, & b^* \leq t_b \end{cases} \quad (4-51)$$

其中,  $t_a$  和  $t_b$  分别为输入图像中  $a^*$  通道和  $b^*$  通道中颜色分布的自适应阈值, 利用颜色直方图计算得出, 用来调整禁止返回机制的时机。计算方法用最大类间方差法来实现。由于模型所关心的是在相应通道上比较显著的区域, 所以可以把相应特征图分成显著和不显著两个部分, 这两者之间的类间方差越大, 就说明这两个部分的差别越大, 也就是说, 得到的显著区域越显著。鉴于模型要找的是最显著的那部分刺激, 因而可以用能够使得类间方差最大的特征值作为阈值, 对低于这一阈值的部分置为 0, 不再处理, 这样可以很好地调节禁止返回的时机, 保证不会将注意焦点指向不显著的区域, 同时还可以大大减少模型的计算量。另外, 由于这一阈值是根据图像的相应特征通道计算得出的, 因而对于不同的图像具有自适应性, 不会因图像的整体上的颜色差异而漏掉最显著的区域。

## 2. 提取显著区域

进一步地, 根据这些亮度通道和颜色通道, 可建立 5 个高斯金字塔  $L^*(\sigma)$ 、 $R^*(\sigma)$ 、 $G^*(\sigma)$ 、 $Y^*(\sigma)$  和  $B^*(\sigma)$ 。若中央周边差操作记为  $\ominus$ , 各通道经多尺度融合后的特征图可由式 (4-52) 至式 (4-56) 求得。

$$L^*(c, s) = |L^*(c) \ominus L^*(s)| \quad (4-52)$$

$$R^*(c, s) = |R^*(c) \ominus R^*(s)| \quad (4-53)$$

$$G^*(c, s) = |G^*(c) \ominus G^*(s)| \quad (4-54)$$

$$Y^*(c, s) = |Y^*(c) \ominus Y^*(s)| \quad (4-55)$$

$$B^*(c, s) = |B^*(c) \ominus B^*(s)| \quad (4-56)$$

其中,  $R^*(c)$  代表图像经尺度为  $c$  的高斯滤波器滤波后相应的红色值;  $R^*(s)$  代表图像经尺度为  $s$  的高斯滤波器滤波后相应的红色值;  $s > c$ 。该公式实际计算的是图像的每个像素点与其周围一定范围内的像素点在某个红色通道上的差值。其他颜色通道及亮度通道类似。

然后将各特征图进一步整合成总特征显著图  $C$ , 如式 (4-57) 所示。

$$C = \bigoplus_c \bigoplus_s \{ N[L^*(c, s)] + N[R^*(c, s)] + N[G^*(c, s)] + N[Y^*(c, s)] + N[B^*(c, s)] \} \quad (4-57)$$

其中,  $\oplus$  表示逐点求和;  $N(\cdot)$  表示归一化算子。求得的  $C$  的最大值点即为当前的注意

点,然后再根据区域生成算法对注意点进行扩展,得到潜在的显著区域。利用该方法得到输入图像的显著区域,在很大程度上降低了下一步特征提取的复杂度,在一定程度上削弱了环境、噪声的影响。

### 3. 后注意阶段

为了进一步识别目标物体,模型对前面提取得到的显著区域进行不变性特征提取,进而与目标物体进行比对,完成目标物体的识别任务。其主要思路如下:首先,计算目标物体标准库的 SIFT 特征,形成识别相应目标物体的模板;然后,计算显著区域的 SIFT 特征,得到能够表征这一区域独特性的不变性特征;最后,通过计算显著区域的 SIFT 特征与标准 SIFT 特征模板的相似度,确定目标物体。

其中的 SIFT 特征是一种局部特征描述子,由 David Lowe 于 1999 年提出,并于 2004 年进行了更深入的发展和完善<sup>[82]</sup>,可以用来处理两幅图像之间发生平移、旋转、仿射变换情况下的匹配问题,具有很强的匹配能力,因而也把这种思路称为 SIFT 特征匹配算法,在物体识别、图像搜索等方面应用十分广泛。而且,在 Mikolajczyk 对包括 SIFT 算子在内的十种局部特征描述子所做的不变性对比实验中,SIFT 及其扩展算法已被证实同类描述子中具有最强的健壮性,对平移、旋转、尺度缩放、亮度变化、遮挡和噪声等具有良好的不变性,对视觉变化、仿射变换也保持一定程度的稳定性,且运算速度相对较快,经优化的 SIFT 匹配算法可以达到实时的要求。

为了更好地理解物体识别的过程,有必要对 SIFT 匹配算法进行简要介绍。SIFT 匹配算法分为两个大的步骤:生成 SIFT 特征和 SIFT 特征向量的匹配。

#### 1) 生成 SIFT 特征

生成 SIFT 特征通过以下步骤来实现。

(1) 检测尺度空间极值点。要想实现所提取的特征具有尺度不变性,就必须构建各种不同尺度下的图像,形成类似金字塔形的图像组。同时,研究人员发现高斯卷积核是能够实现尺度变换的唯一线性核,而且由于计算上的便捷性,使得其成为 SIFT 特征计算过程中当仁不让的选择。因此,构建出的图像金字塔也被称为高斯金字塔。一般情况下,高斯金字塔的下一层图像由上层的图像降采样得到,并利用高斯差分(Difference of Gaussians, DoG)算子实现尺度归一化,使不同层次的图像可以进行像素级的操作,以寻找尺度空间的极值点。这里的尺度空间可以认为是以检测点为中心的立方体,如图 4.25 所示。

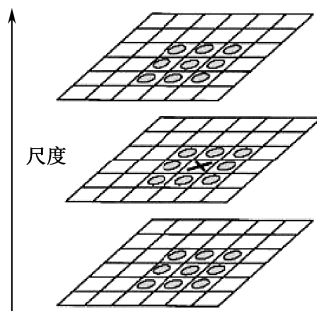


图 4.25 DoG 尺度空间的局部极值检测

其中中间层中间位置的点为检测点(标记为

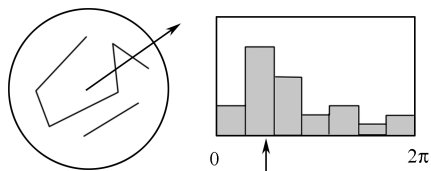


图 4.26 由梯度方向直方图确定主梯度方向

强算法的稳定性。

(3) 为每个关键点指定方向参数。经过第二步的筛选，留下来的点就成为关键点。为了达到旋转不变性，需要为其指定方向参数。采用的方法是：计算关键点周围像素的梯度方向，并统计这些梯度方向的直方图，该关键点的主方向参数就被定为直方图中的最大值（如图 4.26 所示，为了简单，其中的直方图只有 7 个柄）。

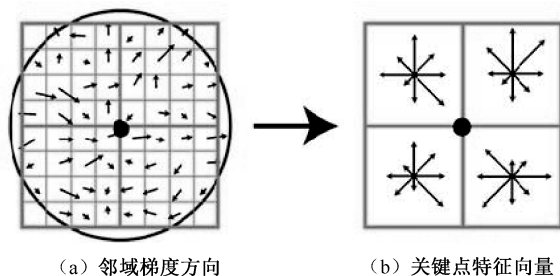
假如在主方向之外，还有一些方向也比较突出，也可以认定为该关键点的辅方向。一个关键点只能有一个主方向，但可以有多个辅方向，这样做的目的是为了增强匹配的鲁棒性。

由此求得的关键点包含以下 3 个信息：位置、所处尺度、方向。此关键点就可以确定一个具有各种不变性的 SIFT 特征区域。

(4) 关键点描述子的生成。描述关键点，就是将其中包含的位置、所处尺度、方向等信息有机融合在一起，为后续的匹配过程提供便利。为了确保旋转不变性，描述时需要将关键点的方向定为坐标轴的方向，然后以关键点为中心取  $8 \times 8$  的窗口。

图 4.27 (a) 中的中央黑点为当前关键点的位置，每个小格表示其周围的一个点，点的梯度方向用箭头方向表示，大小由箭头长度表示，图中圆圈表示一个高斯加权操作（中间的权值大，周围的权值小）。将图中每  $4 \times 4$  的小块作为一个单元，计算 8 个方向的梯度方向直方图，即可形成一个种子点。图 4.27 (a) 中可以形成 4 个种子点，每个种子点有 8 个方向向量信息，如图 4.27 (b) 所示。实际计算过程中，为了增强匹配的稳健性，对每个关键点使用  $4 \times 4$  共 16 个种子点来描述。由于每个种子点有 8

个方向信息，这样对于一个关键点就可以产生  $16 \times 8$  共 128 个数据，即最后形成的 SIFT 特征向量共有 128 维。



(a) 邻域梯度方向

(b) 关键点特征向量

图 4.27 由关键点邻域梯度信息生成特征向量

图 4.28 为一幅图像的 SIFT 特征向量表示。图中共有 19 个 SIFT 关键点，分别用方向和长度各不相同的箭头表示。箭头的起点表示关键点的空间位置，箭头的方向表示关键点的主方向，箭头的



长度表示关键点与周围相邻点的值的差（经过归一化）。

这些关键点对于此目标物体（非机动车道交通标志）具有平移、旋转、尺度缩放等方面的不变性，即不论这个标志在图像中的位置、大小、角度等发生多大的变化，都不会影响这些点的相对位置、所处尺度及向量的方向。因此，求得的 SIFT 特征向量就可以成为一个匹配模板，为下一步匹配操作提供一把尺子。

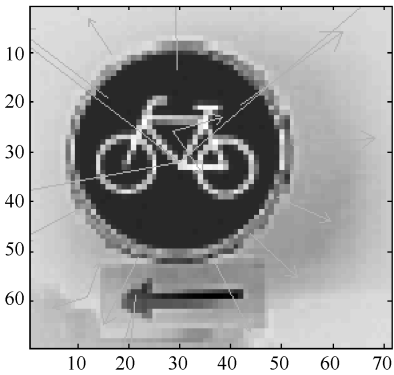


图 4.28 一幅图像的 SIFT 特征向量表示

2) SIFT 特征向量的匹配

在整个模型的识别阶段，首先计算目标物体标准图像库中的图像的 SIFT 特征向量，形成标准 SIFT 特征模板。然后运用视觉注意计算模型寻找可能含有目标物体的显著区域，再用同样的方法，计算出显著区域的 SIFT 特征向量。当这些 SIFT 特征向量生成后，就可以根据比较显著区域的 SIFT 特征向量与标准库中图像的 SIFT 特征向量的相似程度，确定是不是目标物体，完成目标物体的识别。

4.4.2 实验与分析

1. 实验任务

根据对人类视觉系统的生理学研究结果进行分析得知，视觉系统在处理信息过程中，神经节 K 细胞会对强烈的颜色对比产生较强的响应。基于此，本节提出的基于 CIELab 的视觉注意计算模型在颜色空间特征的选择和表示形式上选取了更接近人类视觉认知机制的 CIELab 颜色空间，因而更适用于在视野范围内寻找具有强烈颜色对比的目标物体（如交通标志等），在诸如 ALOI 图像库（其中的图像是在黑色背景下的单一物体）的识别上显示不出明显的优势。

本小节将提出的基于 CIELab 的视觉注意计算模型应用于交通标志检测和识别这一任务，在国内外两个交通标志数据库上对模型性能进行了验证。

近年来，交通安全成为越来越重要的社会问题，因此有关智能汽车安全辅助驾驶系统的研究越来越受到各国研究学者的关注。其中，交通标志的识别是辅助驾驶系统的重要任务之一，分为两个基本技术环节：一是交通标志的检测，即在所观察区域或图片中寻找可能包含交通标志的显著区域；二是交通标志的识别，即在检测过程中得到的显著区域中将交通标志识别出来，给交通参与者提供决策帮助。最早开展交通标志识别研究工作的是于 1986 年开始的高效安全欧洲交通计划（Program for European

Traffic with Highest Efficiency and Unprecedented Safety, PROMETHEUS)<sup>[84]</sup>, 提出了 TSR (Traffic Signs Recognition) 系统。

目前, 交通标志的检测和识别主要有基于颜色的方法和基于标志边缘的方法, 以及将两者相结合的方法<sup>[85-88]</sup>。此类方法源于最经典的图像处理技术, 具有算法成熟、易实现的特点, 在背景比较简单的情況下效果很好, 但当背景比较复杂时, 误检率和错检率会变得很高, 检测和识别效率下降。在现有方法中, 大多采用 RGB 颜色空间, 也有用 HSV 空间的, 基本思想是对采集到的图像进行基于颜色的阈值分割或边缘信息的检测计算, 在此基础上运用神经网络、SVM<sup>[89]</sup>等机器学习方法确定交通标志。

这些方法虽然取得了不错的效果, 但是没有研究交通标志的识别问题的机制, 因此也就不可能从根本上解决交通标志识别问题。其实这个问题本质上是一个视觉认知的问题。为了搞清楚其机制, 我们不妨考察一下人是如何轻易就能发现和认出交通标志的。换句话说, 也就是研究一下为什么人很容易看到路边的交通标志, 而计算机却十分困难。从人类视觉系统抽象出来的视觉注意机制解释了这一问题, 即人类视觉系统对所接收到的信息不是等同对待, 而是只对其中一部分信息感兴趣, 并作详细处理, 大部分信息会被忽略。人们也正是依据视觉注意机制进行交通标志的设计的, 虽然世界各国的交通标志不尽相同, 但总体上是一致的, 例如, 颜色上多用红色、黄色、蓝色等, 形状上多用三角形、圆形、矩形等。这些与周围物体显著不同的设计比较容易引起人们的视觉注意。另外, 当人们驾驶汽车或行走在路上时, 出于安全考虑, 也会有意识地注意寻找交通标志, 这种主观上的注意引导会使交通标志更容易被人类视觉系统发现。

总结世界各地的交通标志, 不难发现, 从交通标志设计目的所体现出的本质来看, 交通标志有两个特点: 一是颜色鲜艳且对比强烈, 与周围环境相比容易引起人的视觉注意; 二是图案标准, 具有很强的规则性, 仿射不变性强。因此, 比较好的方法是运用视觉注意机制来检测交通标志, 找出可能的显著区域, 随后利用其规则性, 运用提取 SIFT 特征的方法来确定显著区域中是否有交通标志, 以及是什么交通标志。

## 2. 实验数据

为了验证提出的模型, 本小节针对部分交通标志牌进行了检测和识别实验。实验数据包括国外和国内两个数据库。国外数据库 (记为数据库 1) 采用荷兰格罗宁根大学 Grigorescu 和 Petkov 设计的交通标志图像数据库<sup>[90]</sup>, 该数据库有 48 幅图像, 分辨率为 360×270 像素, 分为 3 类, 分别对应 3 种交通标志 (非机动车道、人行横道和交叉路口, 如图 4.29 所示), 每类 16 幅。

国内数据库 (记为数据库 2) 由作者在实际路况中用 CCD 相机采集, 共有 120 幅图像, 分辨率为 640×480 像素, 包括三大类交通标志 (禁令、警告与指示标志), 交通标志的标准图采用国家标准《道路交通标志和标线》(GB 5768-2009)<sup>[91]</sup>制作, 如图 4.30 所示。



图 4.29 数据库 1 中的部分标准图



图 4.30 数据库 2 中的部分标准图

### 3. 提取交通标志显著区域

在实验过程中，用基于 CIELab 的视觉注意计算模型对测试图像进行处理，确定潜在交通标志显著区域。实验采用 Matlab R2010a 在 PC (CPU 为 Intel Pentium Dual Core 2.2GHz，内存为 2GB) 上进行仿真，代码参考自 Saliencytoolbox，并根据本节模型进行了相应修改。

图 4.31 至图 4.33 为在数据库 1 上的部分实验结果示例。其中，原模型是指 Itti 提出的视觉注意计算模型，CIELab 模型是指本节提出的基于 CIELab 的视觉注意计算模型。



(a) 原模型得到的显著区域及顺序

(b) CIELab 模型得到的显著区域

图 4.31 原模型和 CIELab 模型在数据库 1 上的实验结果示例 1

在图 4.31 中，原模型在经过 5 次显著点转移后才找到交通标志，而 CIELab 模型首次就能找到交通标志，与原模型相比，在提取交通标志方面表现出更好的效果。在图 4.32 中，原模型提取到的第 2 个显著点才是交通标志，而 CIELab 模型也是首次就能找到交通标志。在图 4.33 中，原模型是第 4 次才找到交通标志，而 CIELab 模型依然是首次就能找到交通标志。

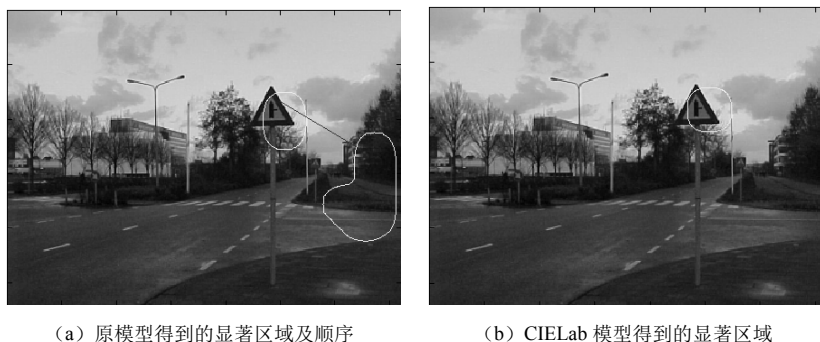


图 4.32 原模型和 CIELab 模型在数据库 1 上的实验结果示例 2



图 4.33 原模型和 CIELab 模型在数据库 1 上的实验结果示例 3

图 4.34 至图 4.36 为在数据库 2 上的部分实验结果示例。在图 4.34 中，原模型提取到的第 2 个显著点是交通标志，而 CIELab 模型首次就能找到交通标志。在图 4.35 中，原模型经过 8 次显著点转移才找到交通标志，而 CIELab 模型提取到的第 2 个显著点便是交通标志。在图 4.36 中，原模型是第 5 次才找到交通标志，而 CIELab 模型在第 3 次转移时找到交通标志，而且找到的交通标志显著区域中交通标志占的区域更大，对于下一步的进一步识别更有利。

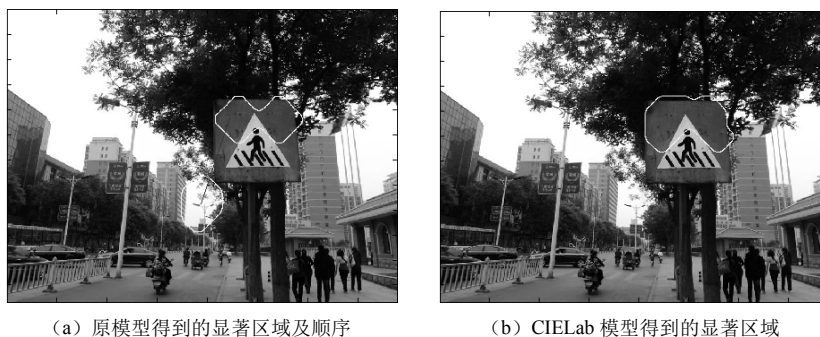


图 4.34 原模型和 CIELab 模型在数据库 2 上的实验结果示例 1

由此可以看出, CIELab 模型比传统的基于 RGB 的模型在提取具有鲜艳颜色的注意点方面具有明显的优势, 这是因为 CIELab 颜色空间的设计更符合人类视觉系统的感知原理, 因而更适用于视觉注意计算模型的特征描述。



图 4.35 原模型和 CIELab 模型在数据库 2 上的实验结果示例 2



图 4.36 原模型和 CIELab 模型在数据库 2 上的实验结果示例 3

#### 4. 识别显著区域中的交通标志

在实验过程中, 对于遴选出的显著区域, 计算其 SIFT 特征, 与事先计算好的标准图像的 SIFT 特征模板进行相似度计算, 给出识别结果。实验采用 Matlab R2010a 在 PC (CPU 为 Intel Pentium Dual Core 2.2GHz, 内存为 2GB) 上进行仿真, 代码参考自 SIFTDemoV4<sup>[92]</sup>, 并根据本节模型进行了相应修改。

图 4.37 至图 4.41 为在数据库 1 上的部分实验结果示例。图 4.37 是数据库 1 中包含某一交通标志 (非机动车道标志) 的一幅图像及用 CIELab 模型提取到的显著区域图。图 4.38 是相应的非机动车道标志标准图及显著区域图的 SIFT 特征向量表示图 (SIFT 特征关键点分别为 7 个和 19 个)。图 4.39 是两者的 SIFT 特征向量匹配示意图 (图中显示匹配到 2 个关键点)。而图 4.40 和图 4.41 分别为数据库 1 中另两个交通标志标准图的 SIFT 特征向量表示图 (SIFT 特征关键点分别为 29 个和 31 个), 以及与

前面提取到的显著区域图的 SIFT 特征向量的匹配示意图（图中均显示没有匹配到关键点）。

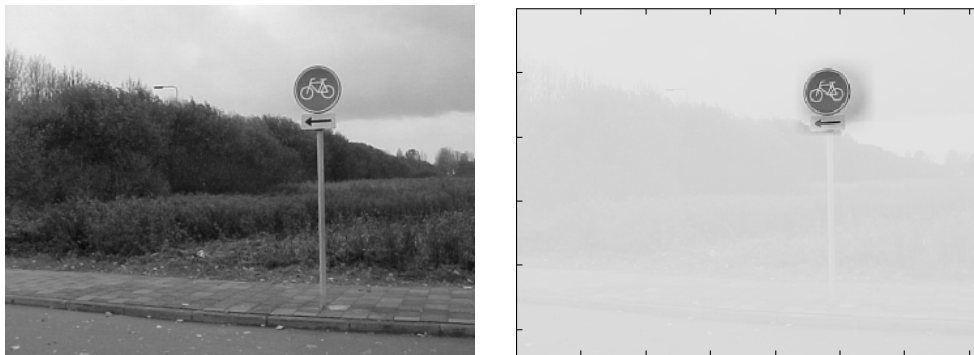


图 4.37 原图及用 CIE Lab 模型提取到的显著区域图（数据库 1）

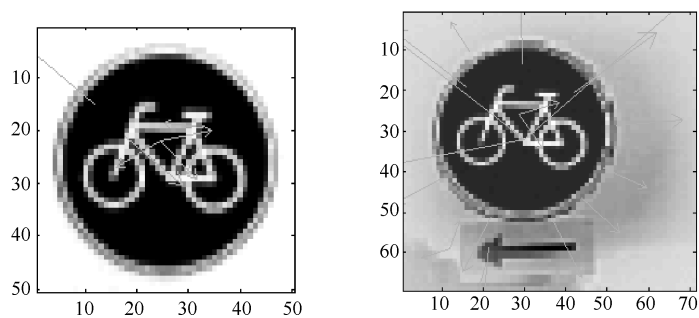


图 4.38 非机动车道标志标准图及显著区域图的 SIFT 特征向量表示图（数据库 1）

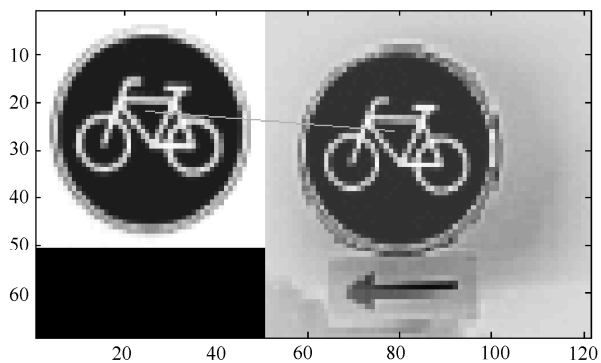


图 4.39 两者的 SIFT 特征向量匹配示意图（匹配到 2 个关键点）（数据库 1）

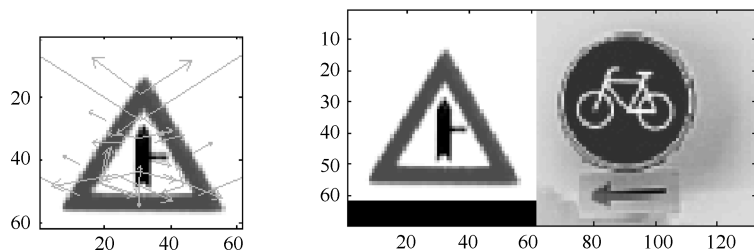


图 4.40 交叉路口标志标准图的 SIFT 特征向量表示图（29 个关键点）及与前面提取到的显著区域图的 SIFT 特征向量的匹配示意图（未找到匹配点）（数据库 1）

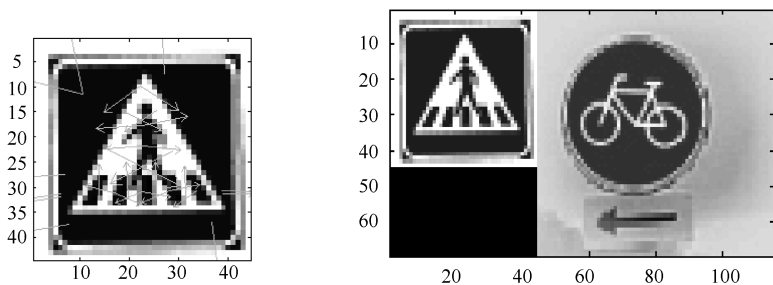


图 4.41 人行横道标志标准图的 SIFT 特征向量表示图（31 个关键点）及与前面提取到的显著区域图的 SIFT 特征向量的匹配示意图（未找到匹配点）（数据库 1）

由此可以得出结论，显著区域是非机动车道标志。

图 4.42 至图 4.46 为在数据库 2 上的部分实验结果示例。图 4.42 是数据库 2 中包含某一交通标志（禁止左转标志）的一幅图像及用 CIE Lab 模型提取到的显著区域图。图 4.43 是相应的禁止左转标志标准图及显著区域图的 SIFT 特征向量表示图（SIFT 特征关键点分别为 53 个和 22 个）。图 4.44 是两者的 SIFT 特征向量匹配示意图（图中显示匹配到 2 个关键点）。而图 4.45 和图 4.46 分别随机选取数据库 2 中另两个交通标志标准图（人行横道标志 1 和人行横道标志 2）作为范例对 SIFT 特征向量匹配过程进行说明，人行横道标志 1 和人行横道标志 2 交通标志标准图的 SIFT 特征关键点分别为 68 个和 57 个，但是从对应的 SIFT 特征向量的匹配示意图中不难发现，两者均没有匹配到关键点，其他的交通标志标准图与此结果相同。



图 4.42 原图及用 CIELab 模型提取到的显著区域图（数据库 2）

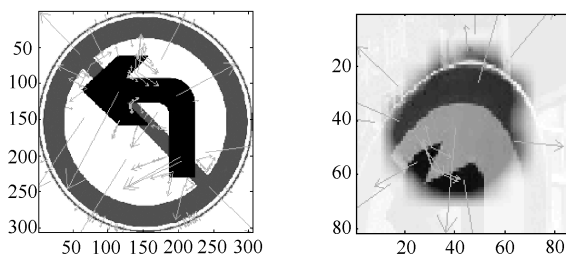


图 4.43 禁止左转标志标准图及显著区域图的 SIFT 特征向量表示图（数据库 2）

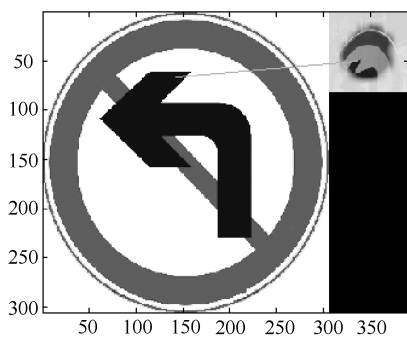


图 4.44 两者的 SIFT 特征向量匹配示意图（匹配到 2 个关键点）（数据库 2）

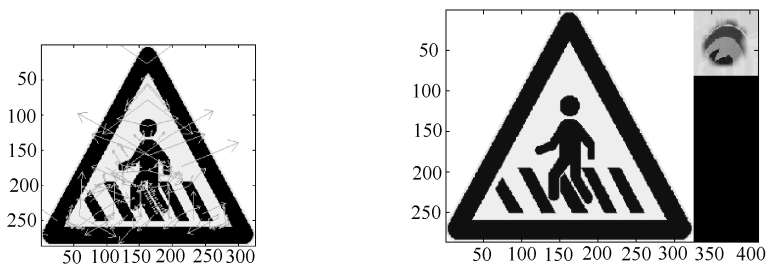


图 4.45 人行横道标志 1 标准图的 SIFT 特征向量表示图（68 个关键点）及与前面提取到的显著区域图的 SIFT 特征向量的匹配示意图（未找到匹配点）（数据库 2）



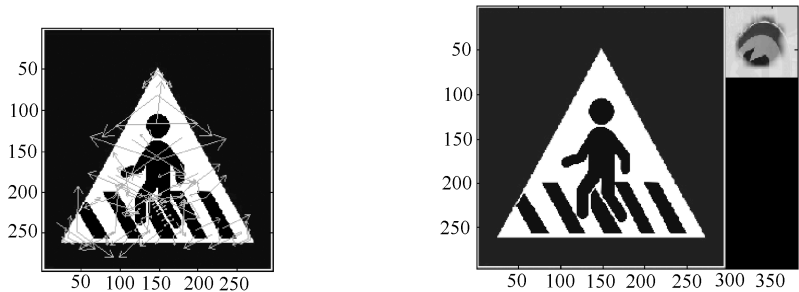


图 4.46 人行横道标志 2 标准图的 SIFT 特征向量表示图（57 个关键点）及与前面提取到的显著区域图的 SIFT 特征向量的匹配示意图（未找到匹配点）（数据库 2）

由此可以得出结论，显著区域是禁止左转标志。

### 5. 实验分析

视觉注意计算模型中的禁止返回机制可以实现注意焦点的转移。如果转移次数太少，可能会漏掉目标物体；相反，如果转移次数太多，则会增加算法的时间复杂度。根据经验，实验设定最大转移次数为 4。实验结果如表 4-1 所示。

表 4-1 模型的识别率、注意焦点平均转移次数












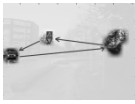















数据库	识别率	注意焦点平均转移次数
1	93.8%	1.92
2	87.0%	1.1

表 4-1 中的注意焦点平均转移次数表示视觉注意计算模型是第几次命中待测交通标志所在区域。每次计算显著区域需要不到 100ms 的时间，而由于 SIFT 特征匹配过程中的区域面积非常小，所以 SIFT 特征匹配过程平均需要不到 30ms。

从表 4-1 中的数据可以看出，本节提出的模型对不同的交通标志识别任务均有很好的效果，不仅识别率高，而且识别效率也不低，可以达到实时检测和识别的效果。另外，数据库 2 的识别率稍低于数据库 1，原因是测试图像中的干扰成分较数据库 1 强烈，若增加转移次数，则可提高识别率，但会付出相应的时间代价。

部分样本的识别结果分析如表 4-2 所示。其中，样本 2 中的注意焦点转移次数为 2，这是因为道路上一辆汽车正在刹车，其红色尾灯亮起，吸引了视觉注意计算模型。样本 4 中有比目标物体更显眼的红色、绿色标志及红色车灯，导致在规定转移次数内没能找到目标物体。样本 6 中左下角的大片绿色刺激了视觉注意计算模型，使模型在第 2 跳才找到目标物体。而样本 9 中的广告牌、LED 广告灯箱、大片树木等复杂且同样具有强烈视觉刺激的干扰源造成了任务的失败。

表 4-2 部分样本的识别结果分析

样本序号	数据库	测试原图	待识别目标	显著区域	注意焦点转移次数	识别结果
1	1				1	正确
2	1				2	正确
3	1				1	正确
4	1				4	错误
5	2				1	正确
6	2				2	正确
7	2				1	正确
8	2				1	正确
9	2				4	错误

## 参 考 文 献

- [1] 维基百科. 视觉系统[EB/OL]. <http://zh.wikipedia.org/zh/视觉系统>.
- [2] F A. Some informational aspects of visual perception[J]. Psychological Review, 1954, 61:183-193.
- [3] Barlow H B. Possible principles underlying the transformation of sensory messages // Rosenblith W A, ed. Sensory Communication[Z]. Cambridge,MA: MIT Press, 1961: 217-234.
- [4] Vinje W E, Gallant J L. Sparse Coding and Decorrelation in Primary Visual Cortex During Natural Vision[J]. Science, 2000,287(5456): 1273-1276.
- [5] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. 1996,381(6583):607-609.
- [6] Olshausen B A, Anderson C H, Van Essen D C. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information [J]. The Journal of Neuroscience, 1993,13(11):4700-4719.
- [7] Field D J. Relations between the statistics of natural images and the response properties of cortical cells[J]. J. Opt. Soc. Am. A, 1987,4:2379-2394.
- [8] Daugman J G. Entropy reduction and decorrelation in visual coding by oriented neural receptive fields[J]. IEEE Transactions on Biomedical Engineering, 1989,36(1):107-114.
- [9] Marr David. 视觉计算理论[M]. 姚国正, 刘磊, 王云九, 译. 北京: 科学出版社, 1988.
- [10] Treisman A M, Gelade G. A Feature-Integration Theory of Attention[J]. Cognitive Psychology, 1980,12(1):97-136.
- [11] Itti L. Models of Bottom-Up and Top-Down Visual Attention[D]. Pasadena, California, 2000.
- [12] Burt P J, Andelson E H. The Laplacian pyramid as a compact image code[J]. IEEE Transactions on Communication, 1983,31(4):532-542.
- [13] 龙甫荟, 郑南宁, 王爱群. 基于非均匀采样及选择注意机制的多分辨率边缘检测[J]. 电子学报,1998(05):97-99.
- [14] Broadbent D E. A mechanical model for human attention and immediate memory[J]. Psychological Review, 1957,64(3):205-215.
- [15] Anne M T. Contextual cues in selective listening[J]. Quarterly Journal of Experimental Psychology, 1960,12(4).

- [16] Treisman A M, Gelade G. A Feature-Integration Theory of Attention[J]. Cognitive Psychology, 1980,12(1):97-136.
- [17] Ulric N. Cognitive psychology[M]. East Norwalk,CT,US:Appleton- Century-Crofts, 1967.
- [18] Deutsch J A, Deutsch D. Attention: Some Theoretical Considerations[J]. Psychological Review, 1963,70(1):80-90.
- [19] Kahneman D. Attention and effort[M]. Englewood Cliffs, 1973.
- [20] Koch C, Ullman S. Shifts in Selective Visual Attention:Towards the Underlying Neural Circuitry[J]. Human Neurobiology, 1985(4):219-227.
- [21] Itti L. Automatic attention-based prioritization of unconstrained video for compression: Human Vision and Electronic Imaging IX, January 19, 2004 - January 21, 2004, San Jose, CA, United States, 2004[C]. SPIE.
- [22] Itti L, Koch C, Niebur E. Model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.
- [23] Itti L, Koch C. Comparison of feature combination strategies for saliency-based visual attention systems[C] // Proceedings of SPIE:The International Society for Optical Engineering, 1999,3644:473-482.
- [24] Itti L, Koch C. Feature combination strategies for saliency-based visual attention systems[J]. Journal of Electronic Imaging, 2001,10(1):161-169.
- [25] Itti L. Real-time high-performance attention focusing in outdoors color video streams: Human Vision and Electronic Imaging VII, January 21, 2002-January 24, 2002, San Jose, CA, United States, 2002[C]. SPIE.
- [26] Itti L. Models of Bottom-Up and Top-Down Visual Attention[D]. Pasadena, California, 2000.
- [27] Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.
- [28] Sun Y, Fisher R. Object-based Visual Attention for Computer Vision[J]. Artificial Intelligence, 2003,146(1):77-123.
- [29] 张鹏, 王润生. 基于视点转移和视区追踪的图像显著区域检测[J]. 软件学报, 2004(06):891-898.
- [30] 张鹏, 王润生. 静态图像中的感兴趣区域检测技术[J]. 中国图像图形学报, 2005(02): 142-148.
- [31] Itti L. Automatic foveation for video compression using a neurobiological model of visual attention[J]. IEEE Transactions on Image Processing, 2004,13(10):1304-1318.

- [32] Itti L, Baldi P. A principled approach to detecting surprising events in video: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, June 20, 2005 - June 25, 2005, San Diego, CA, United States, 2005[C]. Institute of Electrical and Electronics Engineers Computer Society.
- [33] Itti L, Koch C. Feature Combination Strategies for Saliency-Based Visual Attention Systems[J]. Journal of Electronic Imaging, 2001,10(1): 161-169.
- [34] Itti L, Koch C. Comparison of feature combination strategies for saliency-based visual attention systems[C] // Proceedings of SPIE:The International Society for Optical Engineering, 1999,3644:473-482.
- [35] Baluch F, Itti L. Mechanisms of Top-Down Attention[J]. Trends in Neurosciences, 2011,34:210-224.
- [36] Voorhies R C, Elazary L, Itti L. Application of a Bottom-Up Visual Surprise Model for Event Detection in Dynamic Natural Scenes, 2010[C]. May.
- [37] Navalpakkam V, Itti L. An integrated model of top-down and bottom-up attention for optimizing detection speed: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2006, June 17, 2006 - June 22, 2006, New York, NY, United States, 2006[C]. Institute of Electrical and Electronics Engineers Computer Society.
- [38] Carmi R, Itti L. Bottom-up and top-down influences on attentional allocation in natural dynamic scenes, 2004[C]. May.
- [39] 张鹏. 图像信息处理中的选择性注意机制研究[D]. 长沙: 国防科学技术大学, 2004.
- [40] Matthew D B, Jun S. A Neural Network Implementation of a Saliency Map Model[J]. Neural Networks, 2006(19):1467-1474.
- [41] Carota L, Indiveri G, Dante V. A software-hardware selective attention system[J]. Neurocomputing, 2004(58):647-653.
- [42] Park S J, Ban S W, Sang S W. Implementation of Visual Attention System Using Bottom-up Saliency Map Model:ICANN/ICONIP 03, 2003[C].
- [43] Park S J, Shin J K, Lee M. Biologically Inspired Saliency Map Model for Bottom-up Visual Attention[J]. LNCS 2525, 2002:418-426.
- [44] Satoh S, Miyake S. A Model of Overt Visual Attention Based on Scale-Space Theory[J]. Systems and Computers in Japan, 2004,35(10):1-13.
- [45] 张鹏, 王润生. 由底向上视觉注意中的层次性数据竞争[J]. 计算机辅助设计与图形学学报, 2005(8):1667-1672.
- [46] Walther D. Interactions of Visual Attention and Object Recognition: Computational Modeling, Algorithms, and Psychophysics[M]. 2006.

- [47] Walther D, Koch C. Modeling Attention to Salient Proto-Objects[J]. *Neural Networks*, 2006,16(9):1394-1407.
- [48] Craven K M, Downing P E, Kanwisher N. fMRI Evidence for Objects as the Units of Attentional Selection[J]. *Nature*, 1999(401):584-587.
- [49] Yeshurun Y, Kimchi R, Shoua G S, et al. Perceptual Objects Capture Attention[J]. *Vision Research*, 2008.
- [50] Yuanlong Y, Mann G K I, Gosine R G. Modeling of top-down influences on object-based visual attention for robots:Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on, 2009[C]. 19-23 Dec., 2009.
- [51] Ryu G G, Sun I H, Lee S. Covert visual attention by object-based selective visual features and their saliency map: 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV 2009, July 13, 2009 - July 16, 2009, Las Vegas, NV, United States, 2009[C]. CSREA Press.
- [52] Zhang G, Yuan Z, Zheng N, et al. Visual saliency based object tracking: 9th Asian Conference on Computer Vision, ACCV 2009, September 23, 2009 - September 27, 2009, Xi'an, China, 2010[C]. Springer Verlag.
- [53] Kouchaki Z, Nasrabadi A M. A nonlinear feature fusion by variadic neural network in saliency-based visual attention: International Conference on Computer Vision Theory and Applications, VISAPP 2012, February 24, 2012 - February 26, 2012, Rome, Italy, 2012[C]. Inst. for Syst. and Technol. of Inf. Control and Commun.
- [54] Grossberg S, Raizada R. Contrast-Sensitive Perceptual Grouping and Object-based Attention in the Laminar Circuits of Primary Visual Cortex[J]. *Vision Reserach*, 2000(40):1413-1432.
- [55] Sun Y, Fisher R, Wang F. A Computer Vision Model for Visual-Object-Based Attention and Eye Movements[J]. *Computer Vision and Image Understanding*, 2008,112(2): 125-142.
- [56] 曾孝平, 卫立波, 刘国金. 基于图论的视觉注意模型[J]. *四川大学学报(工程科学版)*, 2010(04):125-129.
- [57] Wu T, Gao J, Zhao Q. A Computational Model of Object-based Selective Visual Attention Mechanism in Visual Information Acquisition, 2004[C].
- [58] G U L, M M. Two cortical visual systems[M]. Cambridge, 1982.
- [59] Rybak I A, Gusakova V I, Golovan A V, et al. A model of attention-guided visual perception and recognition[J]. *Vision Research*, 1998,38(15 - 16):2387-2400.
- [60] 田媚. 模拟自顶向下视觉注意机制的感知模型研究[D]. 北京: 北京交通大学, 2007.

- [61] Salah A A, Alpaydin E, Akarun L. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002,24(3):420-425.
- [62] 罗四维. 视觉信息认知计算理论[M]. 北京: 科学出版社, 2010.
- [63] 郑南宁. 认知过程的信息处理和新型人工智能系统[J]. 中国基础科学, 2000(08): 11-20.
- [64] 肖洁. 视觉注意模型及其在目标感知中的应用研究[D]. 武汉: 华中科技大学, 2010.
- [65] 王慧. 空间和目标注意协同工作的视觉注意计算机模型研究[D]. 吉林: 吉林大学, 2010.
- [66] 窦燕, 孔令富. 一种基于视觉注意机制的刀具检测方法[J]. 中国机械工程, 2008 (17):2024-2027.
- [67] 窦燕. 基于空间和物体的视觉注意计算方法及实验研究[D]. 秦皇岛: 燕山大学, 2010.
- [68] 邵静. 协同视觉选择注意计算模型研究[D]. 合肥: 合肥工业大学, 2008.
- [69] 邵静, 高隽. 基于协同感知的视觉选择注意计算模型[J]. 中国图像图形学报, 2008(01):129-136.
- [70] Peters R J, Itti L. Computational mechanisms for gaze direction in interactive visual environments: ETRA 2006-Symposium on Eye Tracking Research and Applications, March 27, 2006-March 29, 2006, San Diego, CA, United States, 2005[C]. Association for Computing Machinery.
- [71] Peters R J, Itti L. Congruence between model and human attention reveals unique signatures of critical visual events: 21st Annual Conference on Neural Information Processing Systems, NIPS 2007, December 3, 2007-December 6, 2007, Vancouver, BC, Canada, 2009[C]. Curran Associates Inc..
- [72] Peters R J, Itti L. Applying computational tools to predict gaze direction in interactive visual environments[J]. ACM Transactions on Applied Perception, 2008,5(2):1-9.
- [73] Peters R J, Itti L. A computational model of task-dependent influences on eye position [C]. 2006, 5. May.
- [74] Xiaodi H, Liqing Z. Dynamic Visual Attention: Searching for coding length increments[J]. 2009.
- [75] 郭海儒. 注意的生成机制与视觉注意计算模型研究[D]. 北京: 北京邮电大学, 2013.
- [76] Elazary L, Itti L. A Bayesian model for efficient visual search and recognition[J]. Vision Research, 2010,50(14):1338-1352.

- [77] Geusebroek J M, Burghouts G J, Smeulders A W M. The Amsterdam Library of Object Images[J]. International Journal of Computer Vision, 2005,61(1):103-112.
- [78] D Walther. Saliencytoolbox[Z]. V2.2 ed. 2007. <http://www.saliencytoolbox.net/index.html>.
- [79] Bishop C M. Pattern Recognition and Machine Learning[M]. Springer, 2006.
- [80] 寿天德. 视觉信息处理的脑机制[M]. 合肥: 中国科学技术大学出版社, 2010.
- [81] 百度百科. 颜色空间[EB/OL]. <http://baike.baidu.com/view/3427413.htm>.
- [82] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.
- [83] 维基百科. Lab 色彩空间[EB/OL]. [http://zh.wikipedia.org/zh/Lab 色彩空间](http://zh.wikipedia.org/zh/Lab%20色彩空间).
- [84] Priese L, Lakmann R, Rehrmann V. Ideogram identification in a real time traffic sign recognition system: Proceedings of the Intelligent Vehicles Symposium '95, Detroit, MI, USA, 1995[C].
- [85] 赵俊梅, 张利平. 基于颜色和形态学的交通标志检测技术的研究[J]. 车辆与动力技术, 2009(4):48-51.
- [86] 何江萍, 马彦. 基于形状信息的三角形交通标志检测方法[J]. 计算机工程, 2010(19):198-199.
- [87] 陈洪波, 王强, 徐晓蓉, 等. 用改进的 Hough 变换检测交通标志图像的直线特征[J]. 光学精密工程, 2009(5):1111-1118.
- [88] Khan J F, Bhuiyan S M A, Adhami R R. Image segmentation and shape analysis for road-sign detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2011,12(1):83-96.
- [89] 朱双东, 刘兰兰. 基于颜色信息与 SVM 网络的交通标志检测[J]. 自动化仪表, 2009(3):69-72.
- [90] Grigorescu C, Petkov N. Distance sets for shape filters and shape recognition[J]. IEEE Transactions on Image Processing, 2003,12(10):1274-1286.
- [91] 中华人民共和国公安部. GB 5768-2009 道路交通标志和标线[S]. 2009.
- [92] David Lowe. SIFT keypoint detector[Z]. V4 ed. 2005. <http://www.cs.ubc.ca/~lowe/keypoints/>.



## 自动图像标注技术

自动图像标注技术旨在根据图像的低层视觉特征，自动推断出图像所表达的语义，并用一个或多个语义关键词对图像语义进行描述。其中，图像特征选择、低层特征到高层语义之间映射模型的建立是自动图像标注技术的两个关键问题，本章主要针对这两个问题展开论述。

### 5.1 概 述

#### 5.1.1 自动图像标注概述及研究意义

虽然 CBIR 技术已经取得了一定成果，但已有的图像检索系统的性能与用户的期望还有一定距离。图像信息检索的根本目的是从大量图像资料中提取用户所需要的图像内容——获得所需要的图像语义信息。为此，首先就要把这些待检索的图像资料的语义信息标注出来，才能从中取出用户的所需。

按照全信息理论<sup>[1,2]</sup>，人们所要获取的信息不像香农信息论的信息那样简单，而是由 3 个相互联系、相辅相成的分量信息所组成的，即语法信息、语义信息、语用信息。其中，语法信息表示的是“事物的状态及其变化方式（也就是事物的状态及其相互关系）”的形式方面；语义信息表示的是“事物的状态及其相互关系”的含义方面；语用信息表示的是“事物的状态及其相互关系”的效用方面。它们所构成的三位一体则称为“全信息”。

虽然人们希望检索的是语义信息，但是，语义信息却存在于“全信息”的三位一体之中，不能独立地存在。因此，当人们研究语义信息问题的时候，不应当孤立地就

语义信息本身来研究语义信息，而应当从语法信息、语义信息、语用信息这个“三位一体”所蕴含的内在联系之中去得到应有的启示。

需要指出的是，在全信息这个三位一体之中，事物的语法信息（形式）是直观的，可以通过人的感觉器官或机器的传感系统直接感知；语用信息（效用）虽然没有语法信息那么直观，但原则上都是可以实际体验到的，因此，可以通过某种体验方式获得；然而，语义信息（含义）却是抽象的，不可能通过感知或体验得到。幸好，抽象的语义信息却可以通过与之相关联的、可以被感知的语法信息和可以被体验的语用信息的联合作用和逻辑演绎来获得。它的基本原理可以参见文献[3]，具体如图 5.1 所示。

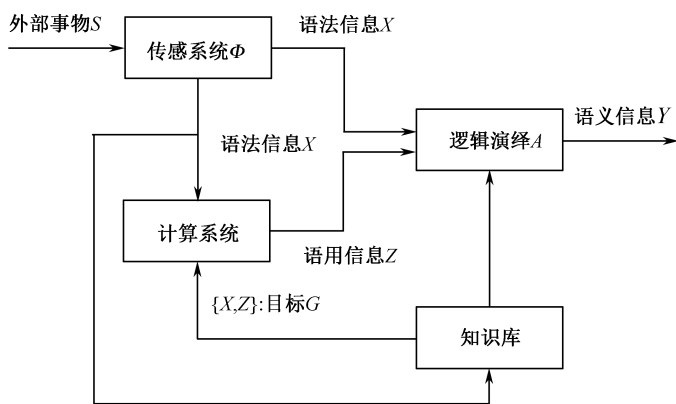


图 5.1 语义信息获取的原理

从图 5.1 中可以看出，语法信息  $X$  通过传感系统得到，语用信息  $Z$  通过在知识库支持下的计算系统得到，而语义信息  $Y$  则通过  $X$  和  $Z$  的联合逻辑演绎得到。其中关于利用传感系统获取语法信息的问题，在数学上表现为映射（理想情况是一对一映射，通常情况是多对一映射），在技术上也已经比较成熟。

关于通过计算系统获得语用信息的问题稍微复杂一些。如图 5.1 所示，给定语法信息  $X$ ，如果知识库事先已经存储了语法信息与相应语用信息的样本对  $\{X, Z\}$ ，就可以直接输出与  $X$  相应的语用信息  $Z$ ；否则，就要根据用户的检索目标  $G$  来计算  $X$  的语用信息。具体的方法是计算语法信息  $X$  与目标  $G$  之间的相关性，相关性为正表示这个语法信息  $X$  是有用的， $X$  越大则效用性越好。

进一步，为了获得语义信息  $Y$ ，可以将已经获得的语法信息  $X$  和相应的语用信息  $Z$  通过逻辑演绎来实现。这是因为所谓语义信息  $Y$ ，在最简单的情况下就是与之相应的语法信息  $X$  和语用信息  $Z$  的逻辑与，即  $Y \leftarrow X \cap Z$ 。例如，什么是“苹果”的语义信息？这是一个抽象的概念，无法直接描述，但可以用与它相应的语法信息（形如圆球、大小如拳、色泽鲜艳）和语用信息（果汁丰富、味道鲜美、有益健康）的逻辑与来表示，即同时具有这样的形式（语法信息）和效用（语用信息）的东西就是苹果。

总之，由于语法信息、语义信息、语用信息构成了“全信息”这个三位一体，因

此理论上就一定可以通过“全信息”这个三位一体的语法信息和语用信息来求得相应的语义信息。

在图像检索中,语法信息就是给定的图像本身(包括它的拓扑形状、颜色分布、纹理结构等要素,它们表现了这个事物的“状态及其相互关系”的形式),基于图 5.1,该问题就变为“如何根据这个基本原理来具体实现由图像的语法信息  $X$  求出相应的语用信息  $Z$ ,进而获得图像的语义信息  $Y$ ”。这可以认为是图像检索和标注的基本理论。

人类对图像的理解并非仅仅建立在客观的图像视觉特征基础之上,采用颜色、纹理、形状等低层特征对图像进行的描述往往与图像本身所包含的语义内容存在较大的差异,即语义鸿沟问题。图像检索的根本目的是让计算机基于语义内容来理解、索引数据库中的图片,即让计算机能够自动理解图片中所包含的语义内容,然后根据图片中的语义内容在图像库中进行相似图片查找,最终实现基于语义的检索<sup>[4,5]</sup>。

描述图像语义最简单、直接的方法就是采用文本表示,即用文本对图像或图像区域的内容进行描述。而且,用文本表示的语义也符合人们的查询习惯。例如,用户想查询包含有“老虎”的图片,如果用户提供一张既包含“老虎”又包含“草”的样例图片,采用 CBIR 方式进行检索,系统可能会返回很多只包含“草”而不包含“老虎”的图片,因为系统根据低层特征进行匹配,无法确切知道用户的查询意图。但是,如果直接采用描述图像内容的文本关键字“老虎”来查询,就不会出现上述结果。相比之下,我们发现采用基于语义内容的图像检索会使检索目标更加明确。如何得到图像的语义描述是基于语义的图像检索的关键技术,因此自动图像标注的研究引起了人们的极大关注<sup>[6,7]</sup>。

由于图像往往具有丰富的语义内容,“百闻不如一见”、“一图值千言”都说明了这个事实,而用户对图像的语义理解又经常表现出主观性和易变性,即不同的用户对同一幅图像的理解与判断经常会不一致,甚至同一个用户在不同的时间或环境下对同一幅图像的语义判断也会不同,因此自动地理解图像的语义内容仍然是一个非常困难的问题。自动图像标注(automatic image annotation)就是给定一幅图像,由计算机系统根据图像的低层特征自动地生成一个或一组可以描述图像语义内容的文本或关键词。从自动图像标注的定义可以看出,图像的视觉特征选择和低层特征到语义标注词映射模型的建立是其两个关键的问题,处理好这两个问题,可以有效地提高自动图像标注的性能,从而提高图像检索系统的准确性。当我们由图像标注技术得到图像相应的语义标签后,就可以根据标注图像的一个或一组语义关键词来实现图像索引,从而,图像检索就可以转化为技术相对成熟的基于文本的图像检索,而基于文本的检索方式具有速度快、效率高的优点,并且其检索精确度也相对较高。

由上述分析可知,图像检索与图像标注密切相关,相互促进。图像检索的迫切需求推动自动图像标注技术的发展,同时,自动图像标注性能的提高也可以更好地提高图像检索效果。实际上,我们可以把自动图像标注视为是基于文本的图像检索的反过程<sup>[8]</sup>,如图 5.2 所示。目前自动图像标注技术已被广泛地应用于多个领域,如医学图

像管理<sup>[9,10]</sup>、人脸识别<sup>[11,12]</sup>和知识产权保护（商标检索）<sup>[13]</sup>等。

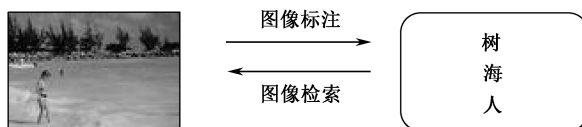


图 5.2 图像标注和图像检索之间的关系

### 5.1.2 自动图像标注的关键问题

随着计算机技术、网络技术和多媒体技术的迅速发展，数字图像的数量呈几何级数增长，这给网络图像检索和浏览带来了巨大的挑战。为了建立快速、准确的图像检索系统，近年来，自动图像标注技术受到了研究者极大的关注，得到了广泛的研究，并被应用到各种领域，如人脸识别、指纹识别、医学图像管理、商标检索等。自动图像标注根据图像的低层视觉特征自动生成对图像语义内容描述的文本标注，由这些文本标注可以实现图像语义索引，从而能够有效提高图像检索的性能。在研究者的努力工作下，自动图像标注技术已经取得了一些令人瞩目的成就，但仍然面临着巨大的挑战，其中，系统由图像低层视觉特征所能获取的信息与人类从图像中得到的信息不一致问题、维数灾难问题、训练数据缺少问题一直是自动图像标注在研究中遇到的难题，主要表现在以下几个方面。

#### 1. 语义鸿沟（semantic gap）问题

语义鸿沟是图像检索和图像语义标注共同面临的一个难题。语义鸿沟是指图像的低层视觉特征并不能有效地描述人类对图像高层语义的认知。例如，颜色、形状及纹理等一系列低层视觉特征并不能有效表达图像的内涵，人们进行图像检索时更关注的是图像的语义内容及图像所表达的意境，也就是图像的高层语义信息。再如，一幅关于节日的图像所表达出的欢乐和喜庆的气氛，仅从图像的视觉特征很难获取到，而需要根据人们日常生活中积累的大量经验和知识来进行推理和判断。正是由于人和计算机对图像相似性的判别依据不同，造成了人所理解的“语义相似”与计算机理解的“视觉相似”之间的语义鸿沟。

#### 2. 维数灾难（curse of dimensionality）问题

维数灾难是图像检索和图像的语义标注领域的另一个关键问题。通常我们需要成百上千的特征来描述一幅图像，但是在低维空间中有效的算法，在高维空间可能变得完全不可行。为了解决这一问题，研究人员开始关注并深入研究各种降维方法，

并将其应用到图像检索和图像标注中。

### 3. 语义概念的相关性和共现性问题

在现实世界中，每幅图像的语义概念都不是独立存在的，它们之间存在着很强的共现模式和语义相关性。传统的学习方法往往假设每个语义类相互独立，因此对这种语义关联性难以很好地把握，很难建模。因此，清晰地学习不同语义概念的边界仍是一个十分困难的问题。

以上这些问题的解决是提高图像标注性能的关键。

## 5.2 图像视觉特征选择

特征描述的就是图像的内容或者说是图像的主要内容。从广义上讲，图像的特征可以是基于文本的特征，如关键字、注释等文本类型的特征；也可以是颜色、纹理、形状等视觉特征，详细内容在第2章已有介绍。在图像处理中，希望提取出来的特征具有尺度不变性、旋转不变性、平移不变性等优良性能，以使图像在发生了缩放、移动或旋转等变化后，不影响标注的效果，或者影响很小。在图像数据分析中，人们常常需要处理具有很多特征且包含大量实例的数据集。在这类数据集中，有些特征是冗余的，甚至是不相关的，冗余特征的存在会降低学习算法的效率，而不相关特征（噪声特征）的存在会有损学习算法的性能。因此，在分析处理这类数据集时，我们有必要对数据进行预处理以去除冗余特征和噪声，即需要进行特征选择。特征选择能给学习算法带来很多好处，例如，可降低其计算代价，可使其生成更易理解的结果和更紧凑、泛化能力更强的模型。

### 5.2.1 视觉特征选择

图像视觉特征选择是图像标注研究最基础的一步。在图像标注问题中，并不是所有的特征都具有很好的区分能力，不可避免地会存在着噪声特征、无关特征和冗余特征，这些特征不仅占用大量的存储空间，而且耗费大量的运算时间。另外，随着应用的不断发展，人们对事物的描述越来越全面，导致描述图像的特征维数也越来越高，数据的高维性不但增加了研究系统的复杂程度，而且还可能严重影响分类的效果，因此，如何选择有效的、区分能力强的特征，有效地构造特征空间是图像标注必须关注的问题。

特征选择方法起始于20世纪70年代，在模式识别领域中，特征选择发挥着非常重要的作用。在有限的样本下，用大量的特征来设计分类器会增加计算开销和降低分

类器的性能。特征选择的目的是去除不相关及冗余的特征，提高数据的分类能力。文献[14]、[15]已经证明了特征选择在剔除不相关特征和冗余特征方面具有良好的性能。

特征选择，就是从现有的特征集  $F = \{f_1, \dots, f_n\}$  中选择一个真子集  $F' = \{f_1, \dots, f_m\}$ ，其中  $n$  为初始特征集的大小， $m$  为特征选择后特征集的大小，要求  $m < n$ 。特征选择并不改变初始特征空间的性质，只是从初始特征空间选择出一部分更有代表性、区分能力更强的特征用于分类。通过穷举式搜索可以得到全局最优解，但是对于数据维数较高的情况，穷举式搜索耗时太大，因此往往在最优化和计算可行性上折中考虑。现有的算法得到的特征子集一般都是局部最优解，并非全局最优解。

特征选择算法主要包括两个步骤：候选特征子集的生成和评价。候选子集的生成过程实际上是一个搜索过程，根据特定的搜索策略得到候选特征子集，然后根据各种评价准则对每个候选子集进行评测<sup>[16,17]</sup>。通常采用不同的启发式搜索算法在特征空间寻找候选子集，如前向序列搜索、后向序列搜索、基于遗传算法的搜索等。根据评价准则的不同，现有的特征选择算法总体可分为3类：过滤器（filter）、封装器（wrapper）和混合器（hybrid）。

近年来，特征选择算法取得了许多研究成果。例如，可采用有监督的方法选择特征，然后根据一定的度量方式进行评价，这类方法有决策树方法<sup>[18]</sup>、粗糙集方法<sup>[19]</sup>、信息熵方法<sup>[20]</sup>、关联规则方法<sup>[21]</sup>等；有根据特征之间的相关性的程度对特征进行选择 and 评价的无监督方法<sup>[22-25]</sup>；还有基于半监督的特征选择方法<sup>[26]</sup>。

特征选择的基本任务是从诸多的特征中去掉不相关及冗余的特征，找出最有效的特征，提高数据的分类能力。在通常的图像标注算法中，一般不考虑特征相对于类别的重要性，都隐含假定图像数据的各维特征在图像标注中所起的作用是一样的。然而，在实际应用中发现，图像低层特征在标注中的重要程度是不一致的，有的特征与某个图像类别相关度高，而有的特征与该类图像的相关度比较低，甚至不相关。举例来说，颜色特征对图像类别为“海”的影响程度较高，而形状特征对该类别图像的影响程度较低，几乎可以不考虑这个特征对这一类别的影响。所以，不能采取“一视同仁”的方法给它们赋予相同的权重，而应合理计算特征与类别的相关度，加大相关度高的特征的权重，减小相关度低的特征的权重，这对避免后续的标注学习算法被弱相关或不相关的特征所支配，具有非常重要的意义。

## 5.2.2 视觉特征加权

特征加权是特征选择的更普遍形式<sup>[27]</sup>。特征选择是从一组特征中挑选出一些最有效的特征达到降低特征空间维数的目的。在特征选择算法中，特征权重的取值要么为0，要么为1，即该类算法是从已有的特征中选择出部分特征用于后续的学习，而没选择到的特征直接被丢弃。特征加权算法则是根据特征相对于分类贡献的不同，为各个

特征赋予不同的权重，相关度高的特征赋予较大的权重，弱相关的特征赋予较小的权重。与特征选择算法不同的是，这里特征权重的取值为  $0 \sim 1$  之间的值，而不单单是只取 0 或 1。也就是说，特征加权算法根据特征的重要程度不同为特征赋予不同的权重，而不是简单地选择或丢弃特征。

近年来，特征加权算法逐渐引起了研究者的重视。文献[28]认为每个特征在分类中所起的作用是不一样的，所以特征的权重应该根据重要程度的不同而有所区别，该文采用基于遗传算法的动态自适应增强算法来进行特征选择，取得了不错的分类效果。文献[29]将图像标注定义为分类问题，每个关键词被视为一个类别，然后利用高斯混合模型来为每个特征进行加权。文献[30]提出了一种特征加权支持向量机的算法，该算法通过计算特征相对于整个分类系统的信息增益来衡量特征的重要性，然后将加权特征应用到支持向量机的分类过程，取得了较高的分类效果。文献[31]根据图像特征分布的直方图分析来确定特征的重要性，然后根据类别的不同，为相对重要的特征赋予较大的权重，最后通过多组实验证明建立本地特征权重是一种有效的特征选择算法。

特征加权旨在根据某种准则为数据集中的各个特征赋予一定的权重，是找出最有效的特征、提高数据的分类能力的一种有效方法。特征加权算法是特征选择算法更加一般的形式。在特征加权中，每个特征被赋予一个  $0 \sim 1$  之间的权重，如果某个特征的相关度比较高，则赋予一个较大的权重；相关度低，则赋予一个较小的权重；不相关则权重为 0。如果特征的权重取值只能是 0 或者 1，则特征加权问题就转化为特征选择问题，所以说，特征选择是特征加权的一种特殊情形。目前，已有很多研究成果是关于应用特征加权提高机器学习算法性能的<sup>[32]</sup>。在这些特征加权算法中，特征权重的计算都是相对整个分类系统的，即它们主要考察特征对整个系统的贡献，而不具体到某个类别，这就使得这些算法比较适合用来作所谓“全局”的特征选择（指所有的类都使用相同的特征权重）。还有一类方法就是针对“本地”的特征选择，即每个类别有自己的特征权重，因为在图像分类中，特征与类别的相关度是不一致的，有的特征，对某个类别有较高的区分度，而对另一个类别则无足轻重。

### 1. 信息增益法

假设集  $D$  为具有类别标号的样本数据， $C = \{c_1, c_2, \dots, c_j, \dots, c_k\}$  为样本的类别标号，共有  $k$  个不同的样本。设  $N_{c_j}$  是训练集  $D$  中  $c_j$  类的样本个数， $N_D$  为数据集  $D$  的样本个数。则对  $D$  中样本正确分类所需的信息为

$$\text{Info}(D) = - \sum_{j=1}^k p_j \log_2(p_j) \quad (5-1)$$

其中， $p_j$  是某个样本属于  $c_j$  的概率，由  $N_{c_j}/N_D$  估计  $p_j$  的取值。假设根据特征  $F$  来对样本集合  $D$  中的数据进行分类，根据训练数据的观测，特征  $F$  具有  $m$  个不同的值  $\{f_1, f_2, \dots, f_i, \dots, f_m\}$ 。利用特征  $F$  可以将  $D$  划分为  $m$  个子集  $D = \{D_1, D_2, \dots, D_i, \dots, D_m\}$ ，

在  $D_i$  中, 每个样本特征  $F$  的取值都为  $f_i$ 。通过对特征  $F$  的划分, 可得特征  $F$  对集合  $D$  进行分类所需要的信息为

$$\text{Info}_F(D) = - \sum_{i=1}^m \frac{N_{c_i}}{N_D} \text{Info}(D_i) \quad (5-2)$$

某个特征的信息增益定义为对集合  $D$  进行分类所需要的信息量与该特征为已知情况下对集合  $D$  进行分类所需要的信息量之差, 即

$$\text{Gain}(F) = \text{Info}(D) - \text{Info}_F(D) \quad (5-3)$$

特征  $F$  的信息增益是由于知道该特征后导致期望信息量减少的信息量。通过上述的方法, 可以计算出集合  $D$  中所有特征的信息增益, 信息增益越大, 表明特征相对于分类的贡献越大, 即具有最高信息增益的特征是特征集中具有最高区分度的特征, 由此可以用信息增益来衡量每个特征相对于分类的重要性。

在信息增益法中, 特征的取值为有限数量的离散值, 所以适合用于文本特征选择及特征可以用有限数字表示的特征权重计算, 如字母图像识别<sup>[30]</sup>。

## 2. Relief-F 算法

Relief 算法<sup>[37]</sup>是 Kira 和 Rendell 在 1992 年提出的, 最初局限于解决两类数据的分类问题, 该算法根据特征分类能力的不同, 为特征赋予不同的权重。Relief 算法从训练集中随机选择一个样本  $x$ , 然后找  $H$  个和  $x$  同一类别的样本、 $M$  个和  $x$  不同类别的样本, 计算其假设间隔, 具体可表示为

$$\theta = \frac{1}{2} [\|x - M(x)\| - \|x - H(x)\|] \quad (5-4)$$

其中,  $H(x)$ 、 $M(x)$  分别表示与  $x$  同类和不同类的最近邻点。假设间隔不仅计算比较简单, 而且其值还可以衡量各维特征的重要程度。在训练集中, Relief 算法通过计算假设间隔的取值, 近似地估计对分类最有用的特征子集, 为特征集中的每个特征计算其相应的特征权重。如果样本点  $x$  与同类样本点在某一维特征上的距离小于  $x$  和不同类样本点的距离, 说明该特征对区分同类和不同类的贡献较大, 则给该特征赋予较大的权重; 反之, 如果  $x$  和同类样本点在某一维特征上的距离大于  $x$  和不同类样本点的距离, 说明该特征对区分同类和不同类的贡献较小, 则给该特征赋予较小的权重。重复进行  $n$  次上述的操作, 然后分别计算各维特征的平均权重。计算得到的特征权重越大, 说明该特征的分类能力越强; 反之, 则说明该特征分类能力越弱。该算法的计算时间随着抽样次数  $n$  和初始特征数  $N$  的增加而线性增加, 所以运行效率较高。

虽然 Relief 算法具有简单、运行效率高、运行效果较好等优点, 但是该算法只能处理两类数据, 因此 1994 年 Kononenko<sup>[38]</sup>扩展了 Relief 算法, 得到的 Relief-F 算法, 可以解决多类问题。在 Relief-F 算法中, 每次从训练样本中随机选择  $k$  个样本, 然后计算样本在各维特征属性上的假设间隔, 并累加起来作为属性的权重, 更新后属性  $j$  的权重可表示为



$$W_j^{i+1} = W_j^i - \text{diff}[j, x, H(x)]/k + \text{diff}[j, x, M(x)]/k \quad (5-5)$$

在 Relief-F 算法中, 如果属性是离散的, 则函数  $\text{diff}()$  定义为

$$\text{diff}(x_k, y_k) = \begin{cases} 0, & x_k \text{ 和 } y_k \text{ 相同} \\ 1, & x_k \text{ 和 } y_k \text{ 不同} \end{cases} \quad (5-6)$$

如果属性是连续的, 则函数  $\text{diff}()$  定义为

$$\text{diff}(x_k, y_k) = \frac{|x_k - y_k|}{\max(p) - \min(p)} \quad (5-7)$$

其中,  $\max(p)$ 、 $\min(p)$  分别是属性  $p$  值的上、下界。

当某一维特征属性对分类贡献较大时, 则表现为在该特征作用下, 同类样本之间的距离比较近, 而非同类样本之间的距离较远, 这时, 大部分样本的假设间隔会较大, 该特征属性的权重自然会较高; 相反, 如果特征属性与分类相关度很低, 那么样本的属性值表现为一系列随机数, 通过大量样本的计算, 其权重将趋向零或较小的数。Relief 系列算法是典型的特征选择算法, 该类算法可以有效地选择出那些对分类贡献较大的特征, 并赋予较高的权重。

### 3. 基于“本地”特征加权方法

信息增益法与 Relief 系列算法在进行特征权重计算时, 主要考察特征对整个系统的贡献, 它们适合用来进行“全局”特征选择, 即所有的类都使用相同的特征权重。在实际的图像分类中, 特征不仅相对于分类系统的重要程度是不一致的, 而且特征与各个类别的相关度也是不一样的, 有的特征与某个类别的相关度较高, 通过该特征可以较好地区分这个类别和其他类别; 有的特征则与该类别的相关度较低, 在分类中所起的作用较小。例如, 颜色特征对图像类别为“海”的影响程度较高, 而形状特征对“海”的影响程度很低, 几乎可以不考虑这个特征对这一类别图像的影响。因此, 我们提出了一种和类别相关的特征权重计算方法。算法根据图像类别不同, 为其对应的特征赋予不同的权重, 即提高与类别强相关特征的权重, 降低弱相关特征的权重, 丢弃不相关的特征。

图 5.3 给出了在图像标注中普遍使用的 Corel 数据集图像示例。从 Corel 数据集中选择 1000 幅图像, 分别属于 10 个 CD-ROM, 每个 Corel CD-ROM 目录下包括 100 幅图像, 表达同一个主题概念。图 5.3 为从每类中取出一幅图像作为样例。这些类别分别是 sun (太阳)、beach (海滩)、building (建筑)、bus (公共汽车)、desert (沙漠)、elephant (大象)、flower (花)、horse (马)、snow mountain (雪山) 和 food (食物)。

通过对图像特征分布的分析, 我们发现图像特征分布具有这样的特性: 在同一个图像类别中, 如果某个特征的统计分布比较密集, 离散程度比较低, 那么这个特征对这个类别是起支配作用的, 是一个重要的特征; 相反, 如果某个特征的统计分布比较分散, 离散程度比较高, 则这个特征对这个类别重要度就低, 如图 5.4 所示。

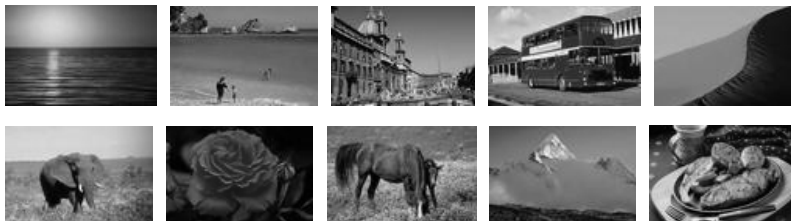
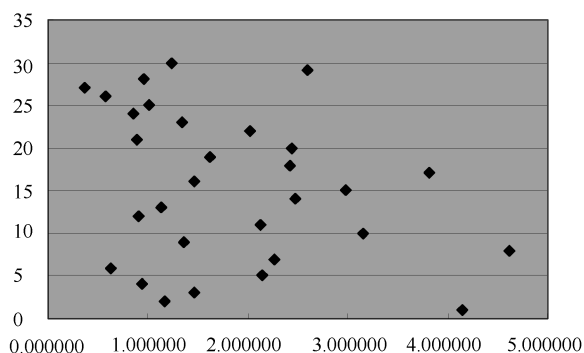
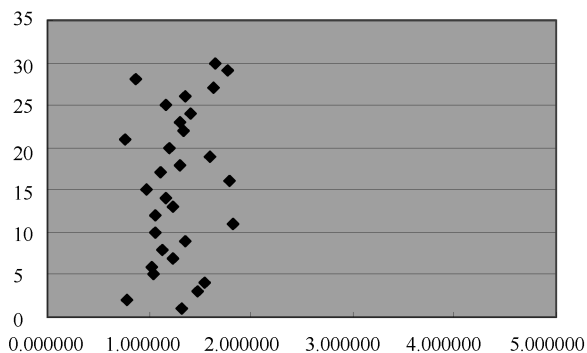


图 5.3 图像示例

图 5.4 (a) 所示为某个图像类别的 30 幅图像的第一维特征分布，图 5.4 (b) 所示为该类别图像的第二维特征分布，可以看出，该类图像的第一维特征分布比较离散，而第二维特征分布则比较集中，可知第一维特征的重要性要低于第二维特征的重要性。如何根据特征的分布特性来计算这个特征的重要程度，是实现特征权重计算的关键。



(a) 第一维特征分布



(b) 第二维特征分布

图 5.4 某类图像前两维特征分布

由于数据的标准差可以很好地反映数据集的离散程度，所以这里采用标准差来衡量图像的特征权重。标准差是各数据偏离平均数的距离（离均差）的平均数，能反映一个数据集的离散程度，可以很好地描述数据的波动性。数据分布越密集，则波动越

小, 相应的标准差也就越小; 相反, 数据分布越分散, 则波动越大, 相应的标准差也就越大。所以, 标准差可以很好地刻画图像特征的分布情况。

在特征加权支持向量机算法中, 每幅图像用一个 17 维的低层特征向量来表示, 有 9 维颜色特征和 8 维纹理特征。由于每类特征的取值范围都不一致, 且相差比较大, 所以需要先对特征进行归一化处理。这里采用下式对特征进行归一化。

$$x'_l = \frac{x_l - m_l}{M_l - m_l} \quad (5-8)$$

其中,  $x_l$  为样本集某个类别中样本的第  $l$  个特征;  $x'_l$  为  $x_l$  经归一化处理后的取值;  $m_l$  为样本集中第  $l$  个特征的最小取值;  $M_l$  为相应的最大取值。假设样本集中第  $j$  类共有  $n_j$  个样本, 每个样本用  $L$  个特征来表示, 其中第  $i$  个样本为  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iL}\}$ , 相应的设该类的特征权重为  $w_i = \{w_{i1}, w_{i2}, \dots, w_{iL}\}$ 。

设第  $j$  类第  $l$  个特征的标准差为

$$\sigma_l = \sqrt{\sum_{i=1}^{n_j} (x_{il} - \bar{x}_l)^2 / (n_j - 1)} \quad (5-9)$$

其中,  $x_{il}$  表示第  $j$  类中第  $i$  个样本的第  $l$  个特征;  $\bar{x}_l$  则表示该类中所有样本第  $l$  个特征的平均值, 即

$$\bar{x}_l = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{il}$$

因为特征取值的分布越密集, 标准差越小, 表示该特征重要程度越高, 相应的权重应该越大, 也就是说特征权重与标准差之间呈反比关系, 所以不直接采用标准差描述图像特征权重。在这里引入一个特征重要度的概念, 特征重要度的取值越大, 表示该特征越重要。

定义特征重要度  $e_l$ ,  $e_l \in [0, 1]$ , 有

$$e_l = \frac{1}{1 + \sigma_l} \quad (5-10)$$

从上式可以看出, 标准差  $\sigma_l$  与特征重要度  $e_l$  成反比关系, 标准差  $\sigma_l$  越小, 特征重要度  $e_l$  越大, 表明该特征越重要。当图像某个特征取值都集中地分布在某一点时, 说明这个特征在该类中起支配作用, 此时  $\sigma_l$  为 0,  $e_l$  取到最大值 1。当  $\sigma_l$  取值较大时,  $e_l$  取值较小, 表明在该类中, 这个特征的贡献度较小。

在具体图像特征权重计算时, 由特征重要度得到特征权重, 第  $j$  类第  $l$  个特征的权重为

$$w_{jl} = e_l / \sum_{l=1}^L e_l \quad (5-11)$$

#### 4. 基于特征加权的支撑向量机图像分类算法

我们将特征加权算法结合支持向量机 (SVM) 算法用于图像分类。文献[39]指出

将图像集正确划分到相应的语义类别将会大大提高基于内容的图像检索系统的性能。目前,很多机器学习算法已被成功用于图像分类,其中,监督机器学习算法是应用较广的一种图像分类方法,是一种有效的降低语义鸿沟的方法<sup>[40]</sup>。在监督学习算法中,支持向量机由于其结构简单、具有全局最优性和较好的泛化能力,被广泛地应用于目标识别、文本分类。

为了更好地提高支持向量机的性能,研究者提出了很多改进算法。文献[41]、[42]提出了基于模糊数学的支持向量机模型(Fuzzy SVM, FSVM)。该模型为了减少噪声和野值的影响,根据样本对分类贡献的不同,为其赋予相应的隶属度,从而提高分类器的性能。由于样本规模大小不同也可能会造成分类性能的降低,文献[43]对不同规模样本赋予不同的权重,该方法根据类别的差异进行了相应的补偿,大大提高了小类别样本的分类精度,该文献对需要重点关注小样本类别精度的应用研究有重要的现实意义。上述算法主要是根据样本的重要性为支持向量机进行加权,并不考虑特征属性对分类模型的影响。我们知道,在 SVM 中,核函数是一个核心问题,核函数的计算常常与样本的特征密切相关,如果样本中存在较多的与分类弱相关或不相关特征,那么,这些特征可能会严重影响核函数的计算,从而最终影响分类器的分类性能。所以,我们提出了基于特征加权算法的支持向量机算法。

在基于特征聚类的图像标注算法中,单纯的图像区域聚类并不能为待标注图像提供语义描述,还必须结合其他算法才能实现自动图像标注。而在基于加权特征支持向量机中,只需用加权特征学习出支持向量机分类模型,就可以将其直接应用于新图像预测。首先利用前文所提的特征权重计算方法进行特征重要性的度量,提高贡献大的特征属性的权重,降低贡献小的特征属性的权重,丢掉不相关的特征,然后将加权特征应用于支持向量机分类中。

在学习过程中,首先需要利用已标注的数据集对 SVM 进行训练学习,然后将训练好的 SVM 应用于测试数据分类。我们用训练集中的样本计算出每个类别的特征权重赋予相应的特征,用加权特征学习 SVM 分类模型。对于待标注图像,分别采用训练样本得到的权重进行加权,然后选用概率最大的作为图像的类别。

基于特征加权支持向量机图像分类的具体算法描述如下。

第 1 步:从每类中随机选取  $M$  个图像作为训练样本。

第 2 步:根据前文的描述,利用式(5-9)至式(5-11)计算训练集每个类中每个特征的权重  $w_{jl}$ 。

第 3 步:将第 2 步得到的权重赋予训练集每个样本,即

$$x'_{il} = w_{jl} x_{il} \quad (5-12)$$

其中,  $w_{jl}$  为第  $j$  类图像第  $l$  个特征的权重;  $x_{il}$  为样本集中该类图像第  $i$  个图像样本的第  $l$  个特征取值;  $x'_{il}$  为  $x_{il}$  更新后的特征权重。

第 4 步:用加权后的特征训练 SVM 分类模型,即

$$\begin{aligned}
& \max \frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j K[\phi(x'_i), \phi(x'_j)] - \sum_{i=1}^l a_i \\
& \text{s.t. } \sum_{i=1}^l a_i y_i = 0 \\
& 0 \leq a_i \leq c, \quad i=1, \dots, l
\end{aligned} \tag{5-13}$$

其中,  $x'_i$  和  $x'_j$  为加权后的特征值。

第5步: 在为未标注图像分类时, 分别用每类特征权重为未标注图像加权, 然后用训练好的 SVM 分类模型分类, 最后选用下式来判别图像类别。

$$\text{label}(B) = \max_{j=1,2,\dots,N} \arg[p(w_j | B)] \tag{5-14}$$

其中,  $p_i(w_j | B)$  表示由 SVM 分类模型将图像  $B$  分为类别  $w_j$  的概率;  $N$  表示标注词的总个数。

## 5.3 自动图像标注模型

建立自动图像标注模型是图像标注技术中的一个最关键问题, 从 20 世纪 90 年代末至今, 研究者提出了诸多经典的标注模型, 如机器翻译模型<sup>[44]</sup>、相关模型<sup>[45]</sup>、隐变量生成模型 LSA<sup>[48]</sup>和 pLSA<sup>[49]</sup>、最大熵模型<sup>[50]</sup>及基于监督学习的分类模型<sup>[51]</sup>等。我们将这些自动图像标注模型分为 3 类: 基于生成模型的标注方法、基于判别模型的标注方法和基于多示例学习的标注方法。基于生成模型的标注方法主要是通过建立标注关键词和图像低层特征的联合概率实现标注; 基于判别模型的标注方法首先将每个标注关键词视为一个类别, 然后使用多类分类技术对待标注图像进行分类; 基于多示例学习的标注方法是将图像标注问题视为多示例学习问题, 从多示例学习的角度来实现图像标注。

### 5.3.1 基于生成模型的标注方法

生成模型方法采用半结构化的平面图结构, 该结构是统计图模型在视觉图像理解领域的拓展和延伸, 是概率统计学和计算机视觉的紧密交叉和融合。生成模型的数据结构复杂, 能表达更多的有用视觉信息和知识信息, 能够很好地完成图像标注的问题。生成模型标注方法的关键是由训练集样本估计标注关键词与视觉区域之间的联合概率。

当进行图像标注时, 相当于通过建模分别得出后验概率分布  $p(\text{word} | \text{image})$ 。生成模型是自顶向下由知识驱动, 分别对标注词先验概率和标注词条件概率密度进行建

模，通过贝叶斯公式将后验概率转换成似然与先验概率的乘积，即

$$p(\text{word} | \text{image}) = p(\text{image} | \text{word}) \cdot p(\text{word})$$

以间接获取目标的后验概率。

在基于图的学习方法中，生成模型是非常重要的一类方法。在自动图像标注中，人们提出了多种基于生成模型的标注方法。下面介绍一系列基于生成模型的图像标注方法。

## 1. 翻译模型

Duygulu 等在 2002 年提出了基于 IBM 翻译模型<sup>[44]</sup> (Translation Model, TM) 的图像标注算法，该算法将图像标注视为一种机器翻译过程，即将图像视觉信息翻译成标注信息。翻译模型如图 5.5 所示。首先，将训练集中的每幅图像进行分割，得到若干个区域；然后用聚类算法对这些区域进行聚类，每一类别用一个 blob 表示，这样每幅图像就可以表示成一系列 blob 组成的特征向量。文献[44]采用传统的机器语言翻译方法来建立图像区域类别与文本词汇之间的对应关系。假设待标注图像可表示为  $b_i = \{b_{i1}, b_{i2}, \dots, b_{im}\}$ ，关键词集合为  $w_i = \{w_{i1}, w_{i2}, \dots, w_{in}\}$ ，则翻译模型可表示为

$$p(w_i | b_i) = \prod_{j=1}^n p(w_{ij} | b_i) \propto \prod_{\{j|w_{ij}=1\}} \sum_{k=1}^m t_{jk} b_{ik} \quad (5-15)$$

其中， $n$  和  $m$  分别为标注关键词和 blob 的个数； $t_{jk}$  为第  $k$  个 blob 翻译为第  $j$  个标注词汇的概率值，在机器翻译模型中每个 blob 只能对应一个标注关键词，而一个标注关键词却可以对应多个 blob。我们的目标是通过寻找  $t_{jk}$ ，使得条件概率  $p(w_i | b_i)$  取值最大，利用 EM 算法可求得最优解。分析发现，翻译模型易受训练集中出现概率高的词影响，词频较高的词比词频较低的词更容易出现在标注结果中，这并不符合实际标注情况。针对这个问题，Kang 等分别于 2004 年和 2005 年提出了两种改进方法：采用规则化翻译过程来消弱词频对标注结果的影响<sup>[52]</sup>和采用对称模型来消弱词频对标注结果的影响<sup>[53]</sup>。

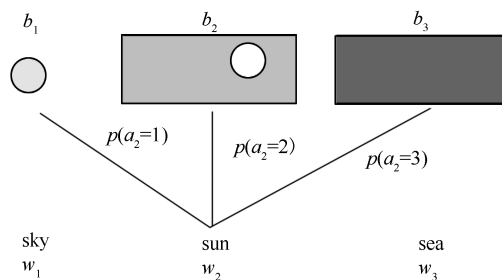


图 5.5 翻译模型

## 2. 跨媒体相关模型

跨媒体相关模型<sup>[45]</sup> (Cross-Media Relevance Model, CMRM) 主要是通过估计标注关键词与图像低层特征的联合概率实现标注的, 是一种很有意义的图像标注模型。在此之后, 研究者在其基础上又提出了很多类似模型, 取得了相对较好的标注效果。

与机器翻译模型一样, CMRM 也用视觉词汇 blob 来表示每一幅图像。由于一般情况下一幅图像包含多个不同的 blob, 而且使用多个词汇进行标注, 共同表达图像的语义。所以, CMRM 假设一组关键词  $W = \{w_1, \dots, w_n\}$  表示的文本信息与一组 blob  $\{b_1, \dots, b_m\}$  表示的视觉对象之间存在着一定的对应关系。给定一幅待标注图像  $I$ , 可以表示为  $I = \{b_1, \dots, b_m\}$ , CMRM 假设存在一个与该图像对应的生成概率密度函数来生成该图像对应的 blob  $\{b_1, \dots, b_m\}$  集合和标注文本  $W = \{w_1, \dots, w_n\}$ 。

CMRM 的目的就是寻找一组最优的标注词集合  $w^*$ , 使得下式的条件概率最大。

$$w^* = \arg \max_{w_i \in W} p(w_i | I) \quad (5-16)$$

其中,  $w_i$  是标注关键词集合  $W$  中的一组词。在 CMRM 中, 利用图像区域  $\{b_1, \dots, b_m\}$  与关键词间的对应关系近似图像与关键词之间的关系, 即

$$p(w_i | I) \approx p(w_i | b_1, b_2, \dots, b_m) \quad (5-17)$$

在图像  $I$  已知的条件下, 有

$$\arg \max p(w_i | b_1, b_2, \dots, b_m) = \arg \max p(w_i, b_1, b_2, \dots, b_m) \quad (5-18)$$

这样, 问题转化为估计联合概率  $p(w_i, b_1, b_2, \dots, b_m)$ 。设已标注样本训练集  $T$ , 待标注图像的联合概率可以写成

$$p(w_i, b_1, b_2, \dots, b_m) = \sum_{J \in T} p(J) P(w_i, b_1, b_2, \dots, b_m | J) \quad (5-19)$$

假设在给定训练图像  $J$  的条件下, 视觉词汇 blob 与标注词之间是相互独立的, 则式 (5-19) 可以表示成

$$p(w_i, b_1, b_2, \dots, b_m) = \sum_{J \in T} p(J) p(w_i | J) p(b_1, b_2, \dots, b_m | J) \quad (5-20)$$

如果假设各个 blob 的观测概率也相对独立, 则有

$$p(w_i, b_1, b_2, \dots, b_m) = \sum_{J \in T} p(J) p(w_i | J) \prod_{i=1}^m p(b_i | J) \quad (5-21)$$

我们要求解  $p(J)$ 、 $p(w_i | J)$  和  $p(b_i | J)$ 。一般假设  $p(J)$  在整个训练集中均匀分布, 对于另外两个条件概率, CMRM 用下面的表达式对它们进行估计。

$$p(w | J) = (1 - a) \frac{\#(w, J)}{|J|} + a \frac{\#(w, T)}{|T|} \quad (5-22)$$

$$p(b | J) = (1 - \beta) \frac{\#(b, J)}{|J|} + \beta \frac{\#(b, T)}{|T|} \quad (5-23)$$

其中,  $\#(w, J)$  和  $\#(b, J)$  分别表示标注词  $w$  或 blob  $b$  出现在图像  $J$  中的次数;  $\#(w, T)$  和  $\#(b, T)$  则分别表示标注词  $w$  或 blob  $b$  出现在训练集  $T$  中的次数;  $|J|$  表示图像  $J$  中所有

的 blob 或者词汇的总数;  $|T|$  表示图像训练集  $T$  中所有的 blob 或者词汇的总数;  $\alpha$ 、 $\beta$  为加权系数。从式 (5-22) 和式 (5-23) 可以看出, 右边的第一项考虑了关键词和 blob 在图像  $J$  上的局部特性, 第二项则考虑了它们在整个训练集上的全局特性。出现两个不同的加权系数  $\alpha$ 、 $\beta$ , 是因为词汇与 blob 有着不同的分布性质。一般来讲, 词汇出现的概率往往具有 zipfian 分布的特性, 即最常用的词汇的频率是第二常用词汇的一倍左右, 后续的大量词汇也遵从这种快速递减性质。而 blob 的分布则相对平稳, 部分原因是系统使用 blob 是通过聚类方法产生的。

### 3. 连续相关模型

在 CMRM 的基础上, Lavrenko 等在 2003 年提出了一种连续相关模型<sup>[46]</sup> (Continuous Relevance Model, CRM), 该模型也是通过估计关键词与图像区域之间的联合概率来求取图像标注的。连续相关模型如图 5.6 所示。CRM 将训练集中的每一幅图像都视为高维矢量空间中的一个点, 然后在这些点上计算某一标注信息在测试图像中的概率。两者不同之处在于 CMRM 是以 blob 来表示图像的, 受聚类算法的影响, 这种离散化方法会导致损失掉一些有用视觉信息; 而 CRM 使用的是直接从图像区域提取的连续特征值, 避免了 CMRM 中的缺点。

设图像中的某个区域  $r$ , 存在一函数  $G$  可以将区域  $r$  映射为一个实特征矢量  $\mathbf{g} \in \mathbf{R}^k$ ,  $\mathbf{g}(r)$  反映了该区域的某些特征或性质。设每一幅图像可以用若干个不相互重叠的区域集合  $r_j = \{r_1, r_2, \dots, r_n\}$  来描述, 该图像文本标注信息为  $w_j = \{w_1, w_2, \dots, w_m\}$ 。CRM 假设图像及其标注生成过程如下<sup>[54]</sup>。

(1) 标注文本  $w_j$  是按照多项式分布  $p_v(*|J)$  独立抽取若干个词  $w_i$  生成的。

(2) 对于任一区域  $r$ , 其对应的特征矢量为  $\mathbf{g}_i$ , 设区域  $r$  是按照概率  $p_r(r_i|\mathbf{g}_i)$  的方式生成的。

(3) 区域描述矢量  $\mathbf{g}_i$  也是按照概率  $p_g(*|J)$  独立同分布的方式生成的。

待标注图像与标注关键词的联合概率为

$$p(w_B, r_A) = \sum_{J \in T} p_T(J) \prod_{b=1}^{n_b} p_v(w_b|J) \prod_{a=1}^{n_a} \int_{\mathbf{R}^k} p_r(r_a|\mathbf{g}_a) p_g(\mathbf{g}_a|J) d\mathbf{g}_a \quad (5-24)$$

其中,  $n_b$  为待标注词汇数目;  $n_a$  为图像区域数目。

我们按下面的方法来估计式 (5-24) 中的几个概率分布。

(1) 按照概率  $p_T(J)$  从训练集  $T$  中选择一幅图像  $J$ , 由于无法确切判断任何一幅图像的出现概率, 所以一般假设其为均匀分布  $p_T(J)=1/|T|$ 。

(2)  $p_r(r_a|\mathbf{g}_a)$  表示的是某个区域的特性, 与具体图像无关。在 CRM 中, 每个图像区域只对应一个生成矢量  $\mathbf{g}_a = \mathbf{g}(r_a)$ , 可以假设  $p_r(r_a|\mathbf{g}_a)$  满足如下分布。

$$p_r(r|\mathbf{g}) = \begin{cases} 1/N_{\mathbf{g}}, & \mathbf{g}(r) = \mathbf{g} \\ 0, & \text{其他} \end{cases} \quad (5-25)$$



其中,  $N_g$  为区域数目。

3) CRM 利用高斯核函数来估计  $p_g(\mathbf{g}|J)$ , 即

$$p_g(\mathbf{g}|J) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2^k \pi^k |\Sigma|}} \exp\{[\mathbf{g} - \mathbf{g}(r_i)]^T \Sigma^{-1} [\mathbf{g} - \mathbf{g}(r_i)]\} \quad (5-26)$$

一般取协方差矩阵  $\Sigma = \beta I$ ,  $I$  为单位矩阵,  $\beta$  描述了核的宽度。

(4)  $p_v(*|J)$  为多项式分布, CRM 利用贝叶斯估计其取值, 即

$$p_v(*|J) = \frac{up_v + N_{v,J}}{u + \sum'_v N'_{v,J}} \quad (5-27)$$

其中,  $p_v$  为标注词  $v$  在训练集中出现的次数;  $N_{v,J}$  为单词  $v$  在图像  $J$  的标注词中出现的次数。

由于连续相关模型不再将区域量化为若干个 blob, 而是直接采用图像区域的特征向量, 从而避免了 CMRM 中的缺点, 所以效果更好。

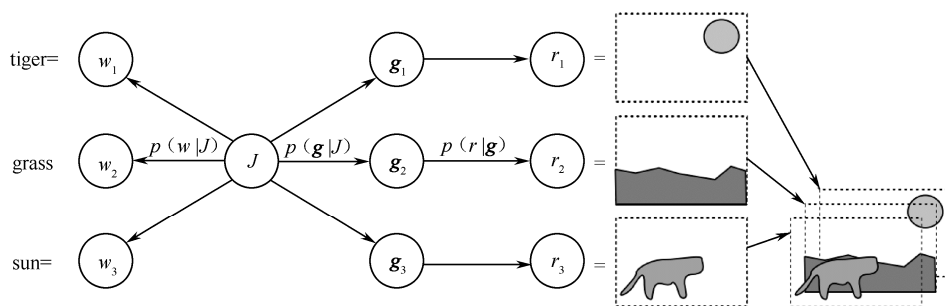


图 5.6 连续相关模型

#### 4. 多伯努利相关模型

在 CRM 和 CMRM 的基础上, 文献[47]提出了改进算法——多伯努利相关模型 (Multiple-Bernoulli Relevance Model, MBRM)。该模型不再选用复杂的图像分割算法, 而采用规则网格划分图像的方法; 另外, 该模型选用多伯努利分布来估计词汇的概率分布, 从而替代了多项式分布。在利用多项式模型进行标注的过程中, 标注词之间存在排斥关系, 过于注重标注关键词的“重要性”, 而在实际图像标注任务中, 关键的问题是每个标注关键词的“存在性”, 即某个标注词是否适合用于描述这幅图像。因此, 在表达词汇分布情况时, 多伯努利分布是一种更加合理的形式。与 CRM 类似, MBRM 仍采用一系列互不重叠的区域集合  $r_j = \{r_1, r_2, \dots, r_n\}$  来表示每幅图像。与 CRM 不同的是, 标注关键词表示为  $w_j = \{0, 1\}^{|V|}$ , 文献[47]按如下方式生成模型。

- (1) 标注文本是按照伯努利分布  $p_v(*|J)$  独立生成的  $|V|$  个词  $w_i$  组成的。
- (2) 对于任一区域  $r$ , 按照概率  $p_g(*|J)$  的方式生成特征矢量  $\mathbf{g}$ 。
- (3) 按照概率  $p_r(*|J)$  的方式生成区域  $r$ 。

所以, 在 MBRM 中, 图像及其标注的联合概率为

$$p(w_B, r_A) = \sum_{J \in T} \left\{ p_T(J) \prod_{a=1}^{n_a} p_g(g_a | J) \prod_{v \in W_B} p_v(v | J) \prod_{v \notin W_B} [1 - p_v(v | J)] \right\} \quad (5-28)$$

对于式 (5-28), 需要估计的参数如下。

(1) 按照概率  $p_T(J)$  从训练集  $T$  中任意选择一幅训练图像, 所以  $p_T(J)$  仍然假设为均匀分布。

(2) 采用与 CRM 同样的方式估计  $p_g(g | J)$  的取值。

设  $p_v(v | J)$  为多伯努利分布, 利用 Bayesian 估计, 可以得到

$$p_v(v | J) = \frac{uN_{v,J} + N_v}{u + N} \quad (5-29)$$

其中,  $N_v$  为标注词  $v$  在训练集中出现的次数;  $N$  为训练集中总的图像数目。

多伯努利相关模型的两个改进较大地提高了模型标注性能。采用规则的矩形来代替图像分割算法, 不仅可以有效地降低计算量, 而且可以显著提高标注性能。另外, 由于采用多伯努利分布估计词汇生成过程是一种更加合理的描述方式, 所以使得标注效果得到进一步提升。

## 5. 基于隐变量的生成模型

潜在语义分析 (Latent Semantic Analysis, LSA) <sup>[48]</sup> 是 Deerwester 等在 1990 年提出的用于进行文本检索的一种方法。该方法认为存在一个潜在的语义空间, 除了那些可观测的数据之外, 不可观测的数据均是存在于这个空间的一些隐变量, 这些隐变量是产生词语和文档的原因。LSA 采用奇异值分解 (Singular Value Decomposition, SVD) 算法将文档从高维空间映射到低维的潜在语义空间。在 LSA 的基础之上, 文献[49]引入概率模型的方式来描述 LSA 的问题, 通常称为 pLSA。LSA 和 pLSA 在文本检索中取得了较好的效果, 2003 年, Monay<sup>[55]</sup>将这两种算法引入到了自动图像标注中。在基于潜在语义的图像标注模型中, 为了能够使用与文本分析类似的模型, 将已标注图像进行分割量化, 表示成一个由文本向量与视觉词汇 blob 连接而成的长向量。由于采用 SVD 的 LSA 模型无法用清晰的概率进行解释, 文献[55]采用 pLSA 对描述图像的混合向量进行分析, 得到如下标注词和图像的联合概率。

$$p(w_j, d_i) = p(d_i) \sum_K P(w_j | z_k) p(z_k | d_i) \quad (5-30)$$

其中,  $z_k$  表示  $K$  个隐变量, 通过 EM 迭代计算, 估算出式中的两个条件概率  $p(w_j | z_k)$  和  $p(z_k | d_i)$ 。在标注过程中, 未标注图像的特征矢量的文本标注信息都用 0 来代替。设待标注图像  $q$ , 推断其标注的后验概率可以表示为

$$p(w_j | q) = \sum_K p(w_j | z_k) p(z_k | q) \quad (5-31)$$

文献[55]的标注模型将视觉信息与文本信息同等对待, 认为它们在构建潜在空间时所起的作用是一致的。然而, 一般情况下, 文本信息与视觉信息重要程度是不一致

的,针对这个问题,文献[56]提出了 pLSA-Words 模型,该模型将两种共现的不同模态特征分别考虑,首先只采用文本特征来定义潜在空间,在标注新图像时,则利用视觉信息  $p(w|d)$  来估计新图像的隐变量条件概率  $p(z|d)$ 。pLSA-Words 方法在对隐变量  $z$  进行初始估计时,并没有加入视觉信息,这样可以较好地避免视觉信息产生的噪声干扰。

其他基于隐变量的图像标注模型还有 2003 年 Blei 等提出的高斯-多项式混合 (Gaussian-Multinomial Mixture, GM-Mixture) 模型<sup>[57]</sup>、高斯-隐 Dirichlet 分配 (Gaussian-Latent Dirichlet Allocation, Gauss-LDA) 模型<sup>[58]</sup>及相关 LDA (Correspondence LDA, Corr-LDA) 模型<sup>[58]</sup>。文献[8]对上述 3 种隐变量算法进行了很好的总结。

近年来,仍有很多工作者致力于图像标注模型的研究。例如,由于 CMRM 的标注效果受聚类算法的影响较大,夏利民等提出了一种基于信息瓶颈算法的图像语义标注方法,该方法首先使用改进的  $k$  均值算法实现图像分割,然后采用信息瓶颈聚类算法对分割后的图像区域进行聚类,以此提高标注模型的性能<sup>[59]</sup>。王梅等提出了一种基于扩展生成语义模型的自动图像标注算法,该算法认为图像初始标注的准确性与图像特征生成的概率估计密切相关,所以首先采用最大权匹配算法估计特征生成的概率,然后利用训练集中关键词之间的相关性提高标注性能<sup>[60]</sup>。这些算法都取得了不错的标注效果。

### 5.3.2 基于判别模型的标注方法

基于判别模型的标注方法将标注关键词视为类别,通过对训练集中已标注语义样本图像的学习,建立分类器,最后在分类器的基础上将未标注的图像进行语义类别的划分。在这些判别方法中,比较典型的判别分类方法有 Boosting 分类方法、基于支持向量机 (Support Vector Machine, SVM) 的图像分类方法和基于神经网络的图像分类算法等<sup>[60]</sup>。图像的标注结果中往往包括多个标注词,所以把图像标注问题看作多分类问题更加确切。本小节主要围绕 Boosting 分类方法和 SVM 分类方法,论述其核心思想和理论方法,并结合图像分类中的相关问题进行数据分析。

#### 1. Boosting 分类方法

Boosting 是一种基于分类器的学习方法,是一种提高任意给定学习算法准确度的方法。它的思想源于 Valiant 提出的 PCA (Probably Approximately Correct) 学习模型。Valiant 和 Kearns 提出了弱学习和强学习的概念,识别错误率小于  $1/2$ ,也即准确率仅比随机猜测略高的学习算法称为弱学习算法;识别准确率很高并能在多项式时间内完成的学习算法称为强学习算法。同时,Valiant 和 Kearns 首次提出了 PCA 学习模型中弱学习算法和强学习算法的等价性问题,即任意给定仅比随机猜测略好的弱学习算法是否可以将其提升为强学习算法。如果二者等价,那么只需找到一个比随机猜测略好

的弱学习算法就可以将其提升为强学习算法，而不必寻找很难获得的强学习算法。

在 Boosting 算法中，每次学习都产生一个假设，称为“弱假设”，最后合并所有弱假设而得到的最终判别函数称为最终假设，记为  $H(x)$ 。定义一个样本为  $(x_i, y_i)$ ，训练集包含样本  $(x_1, y_1), \dots, (x_n, y_n)$ ， $x_i$  为观测值，属于某个实例空间  $X$ ， $y_i$  是  $x_i$  的类别标识，满足  $y_i = f(x_i)$ ， $f$  是学习器要学习的目标函数概念的集合。 $(x_i, y_i)$  是按照某种固定但未知的分布  $D$  随机独立抽取的，样本权值为  $D(1), \dots, D(n)$ ， $D(i) > 0$ ， $1 \leq i \leq n$ ， $\sum_{i=1}^n D(i) = 1$ 。经过一段时间后，学习器输出一个分类假设  $h: X \rightarrow \{0, 1\}$ ，它是对  $f$  的估计。假设空间的集合记为  $H$ ，那么该学习器称为强学习算法，这个强学习算法满足当且仅当对任意小的  $\varepsilon$  和样本分布  $D$ ，此算法都能以概率  $1 - \delta$  [ $\delta \in (0, 0.5)$ ] 输出一个分类假设，它的误差  $\text{error}_D(h) = p[h(x) \neq y] \leq \varepsilon$ 。此外，算法的学习时间必须能表示成  $1/\varepsilon$ 、 $1/\delta$  和其他一些相关参数的多项式形式。经 Boosting 方法提升后得到精度较高的算法就是强学习算法。弱学习算法通常是对一定分布的训练样本给出仅仅强于随机猜测的假设，它与强学习算法满足同样的条件，只是  $\varepsilon \geq 0.5 - \gamma$ ，且常数  $\gamma > 0$ 。Boosting 是一种循环迭代算法，它将一些弱分类器组合为更加复杂的分类器规则，任何 Boosting 实际上都是一个 Boosting 算法和一个弱分类器，最后加权生成一个强分类器。

Boosting 算法的核心思想可以归纳为以下 3 点<sup>[64]</sup>。

(1) 在 Boosting 算法中，样本的权重在没有先验知识的情况下，初始分布为等概率分布，也就是训练集如果有  $n$  个样本，每个样本的分布概率为  $1/n$ 。每循环一次后提高错误样本的分布概率，错分样本在训练集中所占权重增大，使得下一次循环的弱分类器能够集中力量对这些错误样本进行判断。

(2) 准确率越高的弱分类器的权重越大。

(3) 遵循损失函数最小化原则进行迭代循环控制，在强分类器的组合中增加一个加权的弱分类器以提高准确率，减小损失函数，其循环过程就是沿着损失函数的负梯度方向进行最优化的过程。可用累加模型来定义分类器，即

$$H(\mathbf{x}) = a_1 h_1(\mathbf{x}) + a_2 h_2(\mathbf{x}) + \dots + a_t h_t(\mathbf{x})$$

其中， $\mathbf{x}$  是特征向量； $h_t(\mathbf{x})$  是第  $t$  次迭代时得到的弱分类器； $a_t$  是  $h_t(\mathbf{x})$  的权重； $H(\mathbf{x})$  是最终生成的强分类器。Boosting 的本质就是通过最小化指数损失函数来拟合上式所示累加模型。指数损失函数为

$$J(F) = \sum_{i=1}^N e^{-y_i H(\mathbf{x}_i)}$$

其中， $t$  为迭代次数； $y$  是类别标记。通过调整样本的分布  $D(i)$  和选择弱分类器的权重  $a_t$  完成寻优求解。综上所述，弱学习算法生成了一些弱分类器，通过弱分类器对样本的分类结果对样本的分布重新加权，然后产生一系列分类器。对错分样本在下一轮迭代中给予较高的权重，这样新的分类器就关注于那些分错或分类困难的样本。

Boosting 控制训练样本以产生多个假设，通过投票形成分类器集合。Boosting 分

类模型中最具代表性的是 AdaBoost 模型。AdaBoost 算法用具体的学习算法产生弱分类器，并计算弱分类器在训练样本上的错误率，调整训练样本上的概率分布，加大错分样本的权重，在分类器上加权投票建立最终分类器，每个分类器按照其在训练集上的精度再进行加权。AdaBoost 算法有效地解决了早期 Boosting 算法在实际中遇到的困难，最终判别准则的精确度是依赖所有弱学习过程得出的弱假设，因而更能全面地挖掘弱学习算法的能力。AdaBoost 算法执行简单，分类效果较理想，所以受到广泛的关注，之后出现的各种 Boosting 算法基本上都是在 AdaBoost 算法基础上发展起来的。

AdaBoost 算法的基本思想是：首先给出任意一个弱学习算法和训练集  $(x_1, y_1), \dots, (x_n, y_n)$ ，此处  $x_i \in X$ ， $X$  表示某个域或实例空间，在分类问题中是一个带类别标志的集合， $y_i \in Y = \{+1, -1\}$ 。初始化时，AdaBoost 为训练集指定分布为  $1/m$ ，即每个训练例的权重都为  $1/m$ 。接着，调用弱学习算法进行  $T$  次迭代，每次迭代后，按照训练结果更新训练集上的分布，对于训练失败的训练例赋予较大的权重，使得下一次迭代更加关注这些训练例，从而得到一个预测函数序列  $h_1, h_2, \dots, h_t$ ，每个预测函数  $h_t$  也赋予一个权重，预测效果好的，相应的权重越大。 $T$  次迭代之后，在分类问题中最终的预测函数  $H$  采用带权重的投票法产生。单个弱学习器的学习准确率不高，经过运用 Boosting 算法之后，最终结果的准确率将得到提高。

算法初始时每个样本的权重设置为相等的，即  $1/m$ ，经过  $T$  次迭代后，设第  $i$  个训练样本  $(x_i, y_i)$  分布的权重为  $D_t(i)$ ， $D_1(i) = 1/m$ 。每次迭代后，对分类错误的样本加大权重，AdaBoost 算法的描述如下。

输入  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ，其中， $x_i \in X$ ， $y_i \in Y = \{+1, -1\}$ 。

(1) 初始化样本分布  $D_1(i) = 1/m$ ， $i = 1, 2, \dots, m$ 。

(2) For 迭代次数  $t = 1, \dots, T$

① 基于权重分布  $D_t$  执行弱学习算法进行训练。

② 得到弱假设  $h_t : X \rightarrow \{-1, 1\}$ 。

③ 计算弱学习器  $h_t$  的误判率  $\varepsilon_t : p_{i-D_t}[h_t(x_i) \neq y_i]$ ，如果  $\varepsilon_t = 0$  或  $\varepsilon_t \geq 0.5$  则循环结束。

④ 选择  $a_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right) \in \mathbf{R}$  为  $h_t$  的权值，反映其分类准确度。

⑤ 更新权重  $D_{t+1}(i) = \frac{D_t(i) \exp[-a_t y_i h_t(x_i)]}{Z_t}$ ， $i = 1, 2, \dots, m$ 。其中， $Z_t \in \sum_i D_t(i) \exp[-a_t y_i h_t(x_i)]$

是归一化因子，保证  $D_{t+1}$  为概率分布。

For End

(3) 由下式输出最终强学习器。

$$H(\mathbf{x}) = \text{sgn} \left[ \sum_{t=1}^T a_t h_t(\mathbf{x}) \right] \quad (5-32)$$

Boosting 算法因简单高效而受到了人们的很多关注，它使得在实际应用中，不必费力地寻找预测精度很高的算法，而只需找到一个比随机猜测略好的弱学习算法，就

可以通过 Boosting 将其提升为强学习算法，从而也相应地达到提高预测精度的目的。Boosting 算法具有很多优点，它具有较高的正确率，不需要先验知识，只需要选择合适的迭代次数等。但是它速度慢，在一定程度上依赖于训练数据集合和弱学习器的选择，训练数据不充足或者弱学习器太“弱”，都将导致其训练精度的下降。另外，Boosting 易受到噪声的影响，这是因为它在迭代过程中总是给噪声分配较大的权重，使得这些噪声在以后的迭代中受到更多的关注。目前，Boosting 算法仍有许多值得研究的方向，如如何选择合适的迭代次数、如何减少 Boosting 对噪声的敏感等方向。

## 2. SVM 分类方法

支持向量机 (Support Vector Machine, SVM) 是由 Vapnik 等人于 20 世纪 90 年代中期提出的，它是建立在统计学习理论基础上的的一种机器学习方法<sup>[65]</sup>。支持向量机应用 VC 维理论和结构风险最小化原理进行训练，在很大程度上克服了传统机器学习方面所面临的维度高、局部极小值及过学习等困难，并具有良好的推广能力。由于支持向量机出色的学习性能，使得其在模式识别和模式分类等问题中得到广泛的研究与应用<sup>[66]</sup>，近年来一直是机器学习界的研究热点<sup>[67]</sup>。

机器学习的目的是根据给定的训练样本对某系统输入输出依赖关系进行估计，是它能够对未知输出作出尽可能准确的预测。机器学习一般可以表示为变量  $y$  和  $x$  存在一定的未知依赖关系，即遵循某一未知的联合概率  $F(x, y)$  ( $x$  与  $y$  之间的确定性关系可以看作其特例)，机器学习问题就是要基于  $n$  个对立同分布的观测样本，即

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (5-33)$$

其中， $y_i \in \mathbf{R}$ ， $i = 1, 2, \dots, n$ 。在一组函数  $\{f(x, w)\}$  中寻找一个最优的函数  $\{f(x, w_0)\}$ ，对依赖关系进行估计，使期望风险

$$R(w) = \int L[y, f(x, w)] dF(x, y) \quad (5-34)$$

最小。其中， $\{f(x, w)\}$  称作预测函数集， $w \in \Omega$  为函数的广义参数，则  $\{f(x, w)\}$  可以表示任何函数集，通常称作学习函数、学习模型或者学习机器； $L$  是用  $\{f(x, w)\}$  对  $y$  进行预测造成的损失，不同类型的学习问题有不同形式的损失函数。预测函数也称作学习函数、学习模型或者学习机器。

学习的目的就是使期望风险最小，为达到此目的，必须依赖于联合概率  $F(x, y)$ ，但是在实际的机器学习问题中，这一条件是未知的，只能利用已知的训练样本的信息，因此期望风险无法直接计算和进行最小化。为此，在实际的应用中，一般根据大数定律即采用算术平均来代替式 (5-34) 中的实际期望，于是定义了经验风险

$$R_{\text{emp}}(w) = \frac{1}{n} \sum_{i=1}^n L[y_i, f(x_i, w)] \quad (5-35)$$

来逼近式 (5-34) 定义的期望风险。用对参数  $w$  求经验风险  $R_{\text{emp}}(w)$  的最小值来逼近期望风险  $R(w)$  的最小值，这一原则称为经验风险最小化 (Empirical Risk Minimization,

ERM) 原则, 简称 ERM 原则。经验风险最小化原则是目前绝大多数模式识别方法的基础。

在神经网络中, 一开始人们的注意力集中在如何使  $R(w)$  更小。有些情况下, 训练误差过小反而导致推广能力下降, 这就是人们在神经网络研究中往往会遇到的过学习现象, 这也是 ERM 准则不成功的一个典型的例子。出现过学习现象的原因, 一是学习样本不够充分, 二是学习机器的设计不够合理, 这两个问题是相互关联的。总之, 在有限样本的情况下经验风险最小并不意味着期望风险最小, 学习机器的复杂性不但与所研究的系统有关, 而且要和有限的学习样本相适应。

为了研究学习过程一致收敛的速度和推广性, 统计学定义了一系列有关函数集学习性能的指标, 其中最重要的是 VC 维。VC 维是统计学理论的核心内容之一, 它主要反映了函数集的学习能力。这样来定义 VC 维: 对于一个指示函数集, 如果存在  $h$  个样本能够被函数集中的函数按照所有可能的  $2^h$  种形式分开, 则称函数集能够把  $h$  个样本打散, 函数集的 VC 维就是  $h$ , 即函数集能够打散的最大样本数目。如图 5.7 所示, 有 3 个样本, 如果存在一个函数集能够把这 3 个样本划分为 8 种不同的组合方式, 即  $2^3$  种组合, 则函数集的 VC 维为 3。

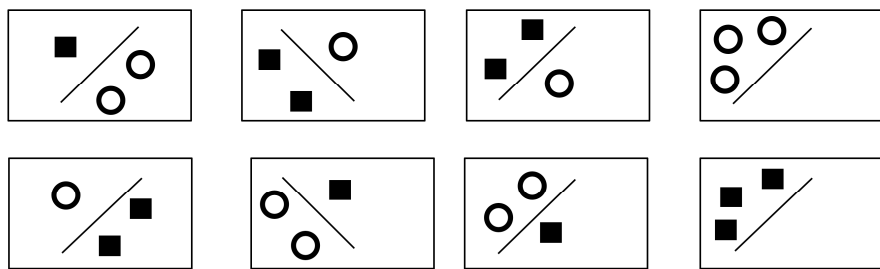


图 5.7 VC 维图示

VC 维是统计学理论中的一个核心内容, 是目前为止对函数性能的最好描述, 反映了函数集的学习能力。VC 维越大, 函数集的学习能力越强, 学习机器也越复杂 (容量越大)。遗憾的是, 目前尚没有通用的关于任意函数集 VC 维的计算理论, 只是对一些特殊的函数集知道其 VC 维。如线性函数的 VC 维等于自由参数的个数,  $n$  维坐标空间线性指示函数集合和线性函数集合的 VC 维是  $n+1$ 。而对于一些复杂的学习机器 (如神经网络), 其 VC 维除了与函数集 (神经网络结构) 有关外, 还受学习算法等的影响, 所以很难确定。

统计学习理论中关于经验风险和真实风险之间关系的重要结论, 称作推广性的界, 它通过下式说明了真实风险  $R(w)$  和经验风险  $R_{\text{emp}}(w)$  之间的关系。

$$R(w) \leq R_{\text{emp}}(w) + \phi(n/h) \quad (5-36)$$

式中,  $\phi(\cdot)$  是置信范围, 是一单调递减的函数;  $h$  是函数集的 VC 维;  $n$  是样本数。上式表明在有限训练样本下, 学习机器的 VC 维  $h$  越大, 复杂性越高, 则置信范围越大, 导致真实风险和经验风险之间可能的差别越大。要取得良好的学习效果, 机器学习过程不但要使经验风险最小, 还要使 VC 维尽可能小, 从而达到缩小置信范围的目的, 最终取得较小的实际风险和较好的推广性能。

传统的学习方法一般采用经验风险来代替真实风险 (期望风险), 但经验风险并不等于真实风险。在追求经验风险最小化的过程中, 往往并不能保证真实风险也最小化, 这就导致了推广能力的下降, 而造成这一现象的原因就在于传统统计学的渐进理论。传统统计学的渐进理论是在样本无穷大的基础上提出来的, 而在实际训练过程中, 样本的个数是有限的, 所以传统的学习方法并不适用于由决策理论导出的期望风险最小化原则。为了解决该问题, 统计学习理论提出了一种新的策略, 即把函数集  $S_k = \{f_w | w \in \Omega\}$  分解为一个函数子集序列或子集结构, 即

$$S_1 \subset S_2 \subset S_3 \subset \cdots \subset S_k \subset S \quad (5-37)$$

使各个函数子集能够按照 VC 维的大小来排列, 也就是要求满足

$$h_1 < h_2 < h_3 < \cdots < h_k < h \quad (5-38)$$

综合考虑经验风险和置信范围使实际风险最小, 这种想法被称为结构风险最小化 (Structural Risk Minimization, SRM) 原则, 其基本思想如下: 要使真实风险  $R(w)$  最小, 只需要使  $R(w) \leq R_{\text{emp}}(w) + \Phi[\frac{m}{h}, \frac{\lg(\eta)}{m}]$  中的  $R_{\text{emp}}(w)$ 、 $\Phi[\frac{m}{h}, \frac{\lg(\eta)}{m}]$  两项相加达到最小; 在要求经验风险  $R_{\text{emp}}(w)$  最小的同时, 也希望学习机模型的泛化能力尽可能大, 即置信范围尽可能小。而该原则在控制经验风险的同时, 也很好地解决了小样本问题。如图 5.8 所示。

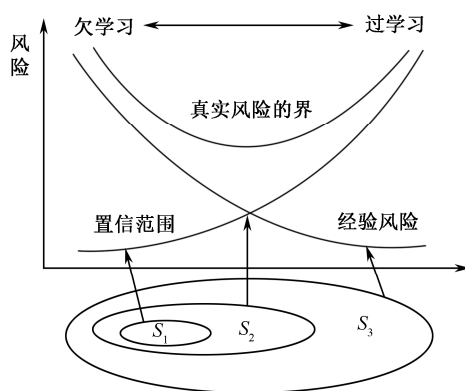


图 5.8 SVM 的经验风险和置信范围

支持向量机的最初研究是从两类线性可分情况发展而来的。给定一个线性可分



的样本集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ ,  $x_i \in \mathbf{R}^n$  是输入向量,  $y_i \in \{+1, -1\}$  是类别标签, SVM 方法就是为了寻找一个两类之间的最优分类面或最优超平面。若超平面  $f(x) = \mathbf{w}^T x + \mathbf{b} = 0$  能将样本正确分为两类, 则最优超平面应使两类样本到超平面最小距离之和最大, 这种最优超平面由离它最近的样本点 (也称支持向量) 决定, 与其他样本点无关, 如图 5.9 所示。

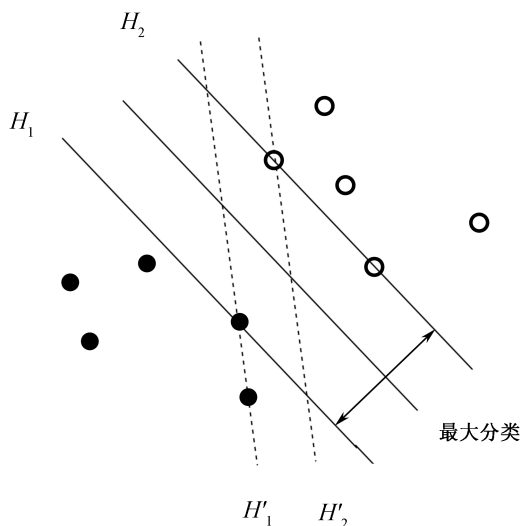


图 5.9 分类间隔和最优分类面示意图

实际中, 有可能存在多个超平面能够实现对样本集正确分类, 考虑图 5.9 所示的两类分类问题, 如  $H_1$ 、 $H_2$  和  $H'_1$ 、 $H'_2$ , 可以看出, 由  $H_1$ 、 $H_2$  得到的分类间隔比沿其他方向得到的分类间隔更大。SVM 的目的就是通过选择使分类间隔最大的分类面, 得到泛化能力最强、结构风险最小的参数模型。支持向量机通过最大化分类间隔来达到这个目的<sup>[71]</sup>。

### 1) 样本线性可分

在样本线性可分的情况下, 可假设存在分类超平面  $\mathbf{w}^T x + \mathbf{b} = 0$ , 为使分类超平面对所有样本正确分类且具备分类间隔, 应满足

$$\begin{cases} \mathbf{w}^T x_i + \mathbf{b} \geq +1, & y_i = +1 \\ \mathbf{w}^T x_i + \mathbf{b} \leq -1, & y_i = -1 \end{cases} \quad (5-39)$$

上式可以写成

$$y_i(\mathbf{w}^T x_i + \mathbf{b}) - 1 \geq 0 \quad (5-40)$$

可以计算出分类间隔

$$\min_{\{x_i|y_i=+1\}} \frac{\mathbf{w}^T x_i + \mathbf{b}}{\|\mathbf{w}\|} - \min_{\{x_i|y_i=-1\}} \frac{\mathbf{w}^T x_i + \mathbf{b}}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (5-41)$$

现在的目标是在满足约束条件下最大化分类间隔  $2/\|\mathbf{w}\|$ ，即要求最小化  $\|\mathbf{w}\|$ ，则求解最优分类超平面问题就可以表示成约束优化问题，即

$$\min \Phi(\mathbf{w}, \mathbf{b}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (5-42)$$

$$\text{s.t. } y_i(\mathbf{w}^T x_i + \mathbf{b}) \geq 1, \quad i = 1, 2, \dots, m \quad (5-43)$$

式 (5-42) 是一个求解不等式约束优化的数学问题，等式成立的点称为支持向量。为了求解该数学优化问题，引入拉格朗日乘子，构造无约束优化问题，其数学表达式如下。

$$\max L(\mathbf{w}, \mathbf{b}, \alpha) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) - \sum_{i=1}^m \alpha_i \{[y_i(\mathbf{w}x_i) + \mathbf{b}] - 1\} \quad (5-44)$$

其中， $\alpha_i \geq 0$  拉格朗日乘子； $\mathbf{b}$  为偏置。

针对式 (5-44)，分别对  $\mathbf{w}$  和  $\mathbf{b}$  求解偏导数，并且令它们形成的等式为 0。通过上述操作，最终可得到 Wolfe 对偶问题，此时式 (5-44) 被转换成一个等式约束的数学优化问题，其数学表达式如下。

$$\max \Phi(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i x_j) \quad (5-45)$$

$$\text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad i = 1, 2, \dots, m \quad (5-46)$$

通过求解，可以得到最优解  $\alpha_i$ 。据此计算

$$\mathbf{w}^* = \sum_{i=1}^m \alpha_i^* y_i x_i \quad (5-47)$$

构建决策函数

$$f(x) = \text{sgn}[(\mathbf{w}^* x) + \mathbf{b}^*] = \text{sgn}\left[\sum_{i=1}^m \alpha_i^* y_i (x_i x) + \mathbf{b}^*\right] \quad (5-48)$$

其中， $\mathbf{b}^*$  可以用任一支持向量来求得。

## 2) 样本线性不可分

当样本集为线性不可分时，需要引入非负松弛变量  $\xi_i$  ( $i = 1, 2, \dots, l$ )，分类超平面的最优问题为

$$\min_{\mathbf{w}, \mathbf{b}, \xi_i} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \quad (5-49)$$

$$\text{s.t. } y_i(\mathbf{w}^T x_i + \mathbf{b}) \geq 1 - \xi_i \quad (5-50)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (5-51)$$

由式(5-49)至式(5-51)有, 当分类出现错误时,  $\xi_i > 0$ , 因此,  $\sum_{i=1}^l \xi_i$  是训练样本中错分样本的上界。这就需要在目标函数中为分类误差分配一个额外的代价函数, 即引入错误惩罚分量。其中,  $C$  为惩罚参数, 它控制对错分样本的惩罚程度,  $C$  越大表示对错分的惩罚越大。采用拉格朗日乘数法求解这个具有线性约束的二次规划问题, 即

$$L = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [y_i (\mathbf{w}^T \mathbf{x}_i + \mathbf{b}) - 1 + \xi_i] - \sum_{i=1}^l \beta_i \xi_i \quad (5-52)$$

其中,  $\alpha_i$ 、 $\beta_i$  为拉格朗日乘子,  $0 \leq \alpha_i$ ,  $0 \leq \beta_i$ , 由此得到

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i = 0 \quad (5-53)$$

$$\frac{\partial L}{\partial \mathbf{b}} = -\sum_{i=1}^l \alpha_i y_i = 0 \quad (5-54)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad (5-55)$$

将式(5-53)至式(5-55)代入式(5-52)中得到对偶最优化问题

$$\max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j \quad (5-56)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \quad (5-57)$$

$$\mathbf{y}^T \boldsymbol{\alpha} = 0 \quad (5-58)$$

最优化求解得到的  $\alpha_i$  中,  $\alpha_i$  可能是: ①  $\alpha_i = 0$ ; ②  $0 < \alpha_i < C$ ; ③  $\alpha_i = C$ 。后两者所对应的样本  $\mathbf{x}_i$  为支持向量 (Support Vectors, SV)。只有支持向量对最优超平面和决策函数有贡献, 支持向量由此得名, 对应学习方法称为支持向量机。在支持向量中, 条件②所对应的样本  $\mathbf{x}_i$  称为标准的支持向量 (Normal Support Vector, NSV); 条件③所对应的  $\mathbf{x}_i$  称为边界支持向量 (Boundary Support Vector, BSV)。根据 KKT 条件, 拉格朗日乘子与约束的积在最优点为 0, 即

$$\begin{cases} \alpha_i [y_i (\mathbf{w} \mathbf{x}_i + \mathbf{b}) - 1 + \xi_i] = 0 \\ \beta_i \xi_i = 0 \end{cases} \quad (5-59)$$

对于标准的支持向量 ( $0 < \alpha_i < C$ ), 由式(5-52)得到  $\beta_i > 0$ , 由式(5-59)得到  $\xi_i = 0$ , 因此, 对于任意标准的支持向量, 满足  $y_i (\mathbf{w} \mathbf{x}_i + \mathbf{b}) = 1$ , 从而计算  $\mathbf{b}$  为

$$\mathbf{b} = y_i - \mathbf{w} \mathbf{x}_i = y_i - \sum \alpha_j y_j \mathbf{x}_j \mathbf{x}_i$$

### 3) 非线性样本线性化处理: 引入核函数

现实中的样本往往是非线性可分的, 为了解决该问题, 支持向量机通过引入核函数, 把输入空间的非线性可分样本映射到一个高维 (以至于无穷维) 的特征空间 (希尔伯特

空间) 中, 此时特征空间中的样本是线性可分的, 从而实现了训练样本的线性可分化, 其原理如图 5.10 所示。

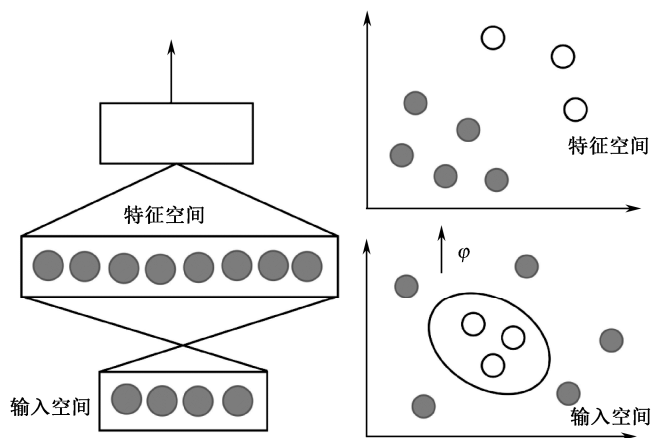


图 5.10 非线性样本线性化处理原理图

在非线情形下, 最优分类超平面为

$$\mathbf{w}\varphi(x) + \mathbf{b} = 0 \quad (5-60)$$

决策函数为

$$f(x) = \text{sgn} [\mathbf{w}\varphi(x) + \mathbf{b}] \quad (5-61)$$

最优分类超平面问题描述为

$$\min_{\mathbf{w}, \mathbf{b}, \xi_i} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \quad (5-62)$$

$$\text{s.t. } y_i [\mathbf{w}^T \varphi(x_i) + \mathbf{b}] \geq 1 - \xi_i \quad (5-63)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, l \quad (5-64)$$

用同样的方法可以得到对偶最优化问题

$$\max \left\{ \begin{aligned} L &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \varphi(x_i) \varphi(x_j) \\ &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \end{aligned} \right\} \quad (5-65)$$

$$\text{s.t. } 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \quad (5-66)$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (5-67)$$

其中,  $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)$  为核函数。参数  $\mathbf{b}$  可由下式计算。

$$\mathbf{b} = \frac{1}{N_{\text{NSV}}} \sum_{x_i \in \text{JN}} \left[ y_i - \sum_{x_j \in J} \alpha_j y_j K(x_j, x_i) \right] \quad (5-68)$$

其中,  $N_{\text{NSV}}$  为标准支持向量数; JN 为标准支持向量的集合;  $J$  为支持向量的集合。

决策函数为

$$f(x) = \text{sgn} \left[ \sum_{i=1}^l y_i \alpha_i K(x_i, x) + b \right] \quad (5-69)$$

在自动图像标注中，待标注的信息往往是若干文本关键词或者有限的类别信息。由于每幅训练图像同时标注有多个关键词，即每个训练样本同时属于多个语义概念类别，则自动图像标注问题是一个典型的多分类学习问题。一种简单而直观地解决多分类学习问题的方法是将其分解为多个独立的二类分类问题来求解。为每个关键词或类别训练对应的二元（binary）分类器，对于待标注图像，只要利用这些分类器来判断每一个关键词或类别是否应该作为该图像的标注，就可以解决图像的标注问题。基于这个思路，在利用支持向量机实现自动图像标注时，通常将该问题视为多分类学习问题，通过将多分类学习问题转化为若干个二类分类学习问题，提出了基于支持向量机的自动图像标注算法。该算法在构造与每个关键词对应的二类分类器时，首先将所有标注该关键词的训练图像作为正例样本，而将所有未标注该关键词的训练图像作为反例样本；然后分别提取正反例图像的低层视觉特征，并据此为该关键词构建 SVM 分类器；最后给定待标注图像，利用每个关键词的分类器实现对其的分类，选择分类标记结果值最高的前几个关键词作为待标注图像的最终标注结果。由于给定的训练图像通常只给出了关键词与图像之间的关联，缺乏关键词与图像中区域的对应关系，即训练图像中存在不属于该关键词语义的区域，而现有的基于多示例学习的自动标注算法考虑到这种标注信息的缺点，因此可得到更加理想的标注效果。

### 5.3.3 基于多示例学习的标注方法

1997 年，Dietterich 等<sup>[73]</sup> 提出了多示例学习框架，该框架主要用来预测药物分子的活性。首先，计算机对已知药物分子进行分析，建立判别模型，然后由模型尽可能正确地判断新分子是否适于制药。已知每个药物分子都有多种低能状态，如果某个分子适合用于制药，那么在它所有可能的低能状态中，至少有一种会和蛋白质分子期望区域耦合得很紧密；而不适合用于制造药物的分子，则不存在这样的低能状态和期望的绑定区域耦合紧密。目前，研究者只知道哪些分子适于制药，并不知道具体是分子中的哪一种状态起到了决定性作用，这使得分子活性预测问题表现出与其他学习问题不同的特点。每个药物分子可能有上百种低能状态，只要其中有一种状态是合适的，这个分子就适于制药。如果直接采用监督学习算法来预测药物分子的活性，即将适于制药的分子的低能状态都视为正例，而将不适于制药的分子的低能状态都作为反例，则会由于正例中噪声示例太多而难以成功预测学习。

Dietterich 等将药物分子活性预测问题抽象为一个新的问题模型：每个分子作为一个包，分子的每种低能状态作为包中的一个示例，该模型称为多示例学习问题模型。

如果该分子适于制药, 则被标注为正, 即表示该分子中存在适于制药的低能状态; 如果某个分子不适于制药, 则被标注为负, 即表示分子中不存在适于制药的低能状态。这就是多示例学习的框架。

多示例学习 (multiple-instance learning) 问题的确切定义可描述如下: 假设训练集由若干个具有概念标注的包 (bag) 组成, 每个包包含若干个示例 (instance), 如果包的标注为正, 则该包中至少有一个示例为正例; 如果包的标注为负, 则包中所有示例的标注都为负。通过对训练集中样本的学习, 建立模型, 用于预测新样本的标注。

20 世纪 90 年代以来, 从例子中学习 (learning from examples) 被认为是最有希望的机器学习途径。如果以训练样本的歧义性 (ambiguity) 作为划分标准, 则目前该领域的研究大致建立在 3 种学习框架 (learning framework) 下, 即监督学习、非监督学习和强化学习。监督学习通过对具有概念标记 (concept label) 的训练例进行学习, 以尽可能正确地对训练集之外的示例的概念标记进行预测。这里所有的训练样本都是有标记的, 因此其歧义性最低。非监督学习通过对没有概念标记的训练例进行学习, 以发现数据中隐藏的结构。这里所有的训练样本都是没有标记的, 因此其歧义性最高。强化学习通过对没有概念标记但与一个延迟奖赏或效用 (可视为延迟的概念标记) 相关联的训练例进行学习, 以获得某种从状态到行动的映射。这里所有的训练样本都是有标记的, 但与监督学习不同的是, 标记是延迟的, 因此强化学习的歧义性介于监督学习与非监督学习之间。

多示例学习框架既不同于监督学习和非监督学习, 也不同于强化学习, 它是一种新的学习框架。与监督学习相比, 多示例学习中的训练示例是没有概念标记的, 这与监督学习中所有训练示例都有概念标记不同; 与非监督学习相比, 多示例学习中的训练包是有概念标记的, 这与非监督学习中的训练样本没有任何概念标记也不同; 而与强化学习相比, 多示例学习中又没有时效延迟的概念。更重要的是, 在以往的各种学习框架中, 一个样本就是一个示例, 即样本和示例是一一对应关系; 而在多示例学习中, 一个样本 (即包) 包含了多个示例, 即样本和示例是一对多的对应关系。因此, 多示例学习中训练样本的歧义性与监督学习、非监督学习、强化学习的歧义性都完全不同, 这就使得以往的学习方法难以很好地解决此类问题。由于多示例学习具有独特的性质和广泛的应用前景, 属于以往机器学习研究的一个盲区, 因此在国际机器学习界引起了极大的反响, 被认为是一种新的学习框架。

多示例学习问世以来得到研究者极大的关注, 并在短短几年时间内取得了一系列引人瞩目的理论成果和应用成果, 被认为是与非监督学习、监督学习和强化学习并列的第四种机器学习框架。多示例学习主要用于药物分子预测<sup>[74]</sup>、股票选择<sup>[79]</sup>、图像检索<sup>[80]</sup>和文本分类<sup>[83]</sup>等领域。常用的方法主要有轴-平行矩形 (axis-parallel rectangles) 算法<sup>[73]</sup>、多样性密度 (Diverse Density, DD) 算法<sup>[78]</sup>、EM-DD 算法<sup>[85]</sup>、MI-SVM 算法和 mi-SVM 算法<sup>[86]</sup>等。本书主要是研究多示例学习在图像标注中的应用。下面介绍几种基于多示例学习的图像标注算法。

### 1. 多样性密度算法

多样性密度 (Diverse Density, DD) 算法是由 Maron 和 Lozano-Perez 在 1998 年提出的, 是解决多示例学习问题的一个典型框架。多样性密度算法的基本思想是在属性空间中寻找一个概念点, 该点附近出现的正包数越多, 负包示例出现得越远, 则该点的多样性密度越大, 即为用户感兴趣概念的概率就越大。

设  $B = \{B_1^+, \dots, B_{p^+}^+, B_1^-, \dots, B_{p^-}^-\}$  为训练包。其中,  $B_i^+$  代表训练集中第  $i$  个正包;  $B_{ij}^+$  代表第  $i$  个正包的第  $j$  个示例;  $B_{ijk}^+$  代表第  $i$  个正包的第  $j$  个示例的第  $k$  个属性的值。同理, 令  $B_i^-$ 、 $B_{ij}^-$ 、 $B_{ijk}^-$  分别代表第  $i$  个负包、第  $i$  个负包的第  $j$  个示例及第  $i$  个负包的第  $j$  个示例的第  $k$  个属性的值。设  $t$  为属性空间中多样性密度最大的点, 可以通过在特征空间最大化  $\Pr(x=t | B_1^+, \dots, B_{p^+}^+, B_1^-, \dots, B_{p^-}^-)$  来确定  $t$  的取值。根据贝叶斯定理, 可以通过最大化似然函数  $\Pr(B_1^+, \dots, B_{p^+}^+, B_1^-, \dots, B_{p^-}^- | x=t)$  来求得多样性密度最大的点。假设, 已知目标概念点  $t$  时, 训练集中各个包之间是条件独立的, 则可以通过下面的 DD 函数来寻找属性空间多样性密度最大的点。

$$DD(x) = \arg \max_x \prod_i \Pr(B_i^+ | x=t) \prod_i \Pr(B_i^- | x=t) \quad (5-70)$$

式 (5-70) 为最大多样性密度函数的定义, 可以选择采用以下的 noisy-or 模型对式 (5-70) 中的乘积项进行例化, 即

$$\Pr(x=t | B_i^+) = 1 - \prod_j [1 - \Pr(x=t | B_{ij}^+)] \quad (5-71)$$

$$\Pr(x=t | B_i^-) = \prod_j [1 - \Pr(x=t | B_{ij}^-)] \quad (5-72)$$

用示例与目标概念点之间的距离来定义该示例与潜在目标概念点之间的概率, 则有

$$\Pr(x=t | B_{ij}) = \exp(-\|B_{ij} - x\|^2) \quad (5-73)$$

直观地考虑, 如果正包中的某个示例距离  $x$  比较近, 并且所有负包都远离  $x$ , 在特征空间  $x$  点的多样性密度会比较大。

下面以图 5.11 为例来理解多样性密度算法。

图 5.11 中的正包和负包示例满足同样的分布, 数字表示正示例, 小圆点表示负示例, 中间的小方块中包含了每个正包中的一个示例, 并且不包含负示例。我们的目标就是找到中间的小方块, 它表示的是多样性密度最大的点。示例的远离程度主要采用欧氏距离衡量。在特征空间有些特征可能是不相关的, 而有些特征的重要性要高于其他特征, 采用下面的公式来计算示例间的欧氏距离。

$$\|B_{ij} - x\|^2 = \sum_k w_k (B_{ijk} - x_k)^2 \quad (5-74)$$

其中,  $x$  为示例空间的点;  $w_k$  为属性的权值,  $w_k$  是一个非负的参数, 是一个可以衡量第  $k$  个属性相关度的加权系数。所以, 对式 (5-70) 的求解, 不仅可以确定多样性

密度最大的点, 还可以得到一组衡量属性相关程度的权重。由于 DD 函数是连续的且是高度非线性的, 所以多样性密度空间中存在多个局部极小点, 通过最大化 DD 函数来寻找目标概念点, 就成为一个优化问题, 普通的 DD 算法一般采用梯度下降法来寻找目标概念点。为了得到最大的多样性密度点, 普通的 DD 算法是以所有的正包中的示例为出发点进行寻优的, 所以 DD 算法的运算开销很大。

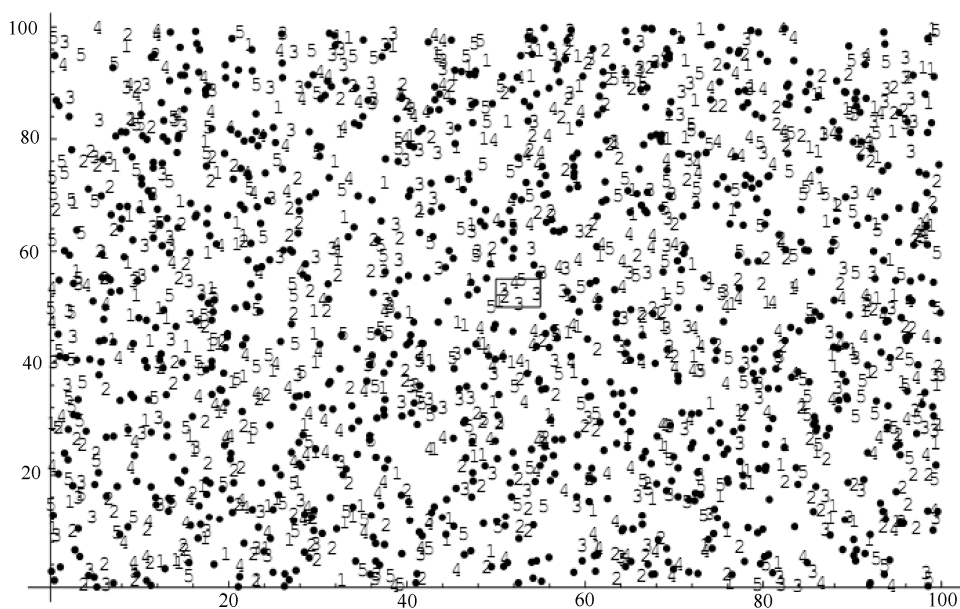


图 5.11 正包和负包中的示例分布

为了扩展 DD 算法的应用范围, Dooly 等<sup>[87]</sup>在 DD 算法的基础上提出了实数输出算法。设给定的示例包的集合为  $B = \{B_1, \dots, B_m\}$ , 标注的集合为  $L = \{l_1, \dots, l_m\}$ , 即  $D = \{(B_1, l_1), \dots, (B_m, l_m)\}$ 。假设  $B_i = \{B_{i1}, \dots, B_{im}\}$ , 其中  $B_{ij}$  表示包  $B_i$  的第  $j$  个示例。设包  $B_{ij}$  中每个示例的标注为  $l_{i1}, \dots, l_{im}$ , 如果是二值标注  $l_{ij} \in \{0, 1\}$ ,  $l_i = l_{i1} \vee l_{i2} \vee \dots \vee l_{im}$ , 求解最大多样性密度点的方法就如前面的基本 DD 算法所述; 如果是实数标注,  $l_{ij} \in [0, 1]$ ,  $l_i = \max \{l_{i1}, l_{i2}, \dots, l_{im}\}$ , 即包的标注是由包中示例标注取值最大的那个决定。在该算法中求解多样性密度点的方法可以通过下面的式子来实现。

$$\arg \max_t \Pr(t | B) = \arg \max_t \prod_i \Pr(t | B_i) \quad (5-75)$$

$$\Pr(t | B_i) = 1 - |l_i - \text{Label}(B_i | t)| \quad (5-76)$$

$$\text{Label}(B_i | t) = \max_j \left\{ \exp \left[ - \sum_k s_k (B_{ijk} - t_k)^2 \right] \right\} \quad (5-77)$$



## 2. EM-DD 算法

DD 算法对后来的研究具有很大的影响,被广泛地应用于各种多示例学习问题中。例如,Maron<sup>[78]</sup>将其用于药物分子活性预测;文献[88]将其用于自然场景分类;文献[89]用 DD 算法实现图像的检索。

研究者根据 DD 算法的思想,提出了新的扩展算法。2002 年,Zhang<sup>[85]</sup>等将 EM 算法与 DD 算法相结合提出了 EM-DD 算法。在多示例学习中,包的标注是由包中最有可能为正的那个示例决定的,即由包中所有示例的概率最高的那个示例决定的,困难的是无法知道包中哪个示例为正的可能性最大。EM-DD 算法的基本思想是将能够决定包的标注的示例视为缺失信息,然后再用 EM 算法来进行估计。算法首先假设已知一个初始点  $h$ ; 在算法的 E 步,从训练包中选出最靠近  $h$  的示例组成一个集合;然后在 M 步,对这些示例使用梯度搜索法估算出一个新的使多样性密度最大的目标点  $h'$ ,并替代  $h$ ;反复进行 E 步和 M 步直到收敛为止。由于不需要像 DD 算法一样,以所有正包示例为起始点进行梯度搜索,所以大大减少了运算时间。其他将多样性密度算法进行扩展的算法参见文献[90]、[91]。

## 3. 基于支持向量机的多示例学习算法

Andrews 等于 2002 年提出了两种基于支持向量机的多示例学习算法: MI-SVM 和 mi-SVM。这两种算法根据多示例学习问题的特性,将支持向量机进行扩展,在使得分类间隔最大的情况下,使其可以解决分类问题。其中,MI-SVM 主要是从包的角度进行分析的,可以实现包的分类;而 mi-SVM 主要是从示例的层面进行考虑的,可以实现示例的分类。

在标准的支持向量机算法中,我们希望找到一个分类超平面把两类样本点分开。存在不同的分类超平面可以把两类样本正确划分,支持向量机以结构风险最小为目标,选择具有最大分类间隔的分类面。文献[65]已经从理论上证明分类间隔越大,分类超平面集合的 VC 维越小。因此,使分类间隔最大的分类面即是泛化能力最强的分类面。MI-SVM 和 mi-SVM 是通过对支持向量机方法进行改进和拓展来解决多示例学习问题的两种典型算法。MI-SVM 是从包的数据层次来讨论图像分类问题的,它的目标是最大化图像包的分类间隔;而 mi-SVM 是从示例的角度考虑分类问题的,将包中的每个示例作为分类对象,其目的是最大化示例样本的分类间隔。

在模式分类问题中,已知条件一般为训练数据及其相应的类别标记  $(x_i, y_i) \in \mathbf{R}^d$ , 训练的目的在于学习得到一个关于模式与标注的分类器函数  $f: \mathbf{R}^d \rightarrow \gamma$ 。对于二分类问题,  $\gamma = \{-1, 1\}$ 。设训练集图像包  $B_1, \dots, B_m$  的标注是已知的,包中示例  $x_1, \dots, x_n$  的标注是未知的,  $B_I = \{x_i | i \in I\}$ ,  $I \subseteq \{1, \dots, n\}$ 。对应每个图像包  $B_I$  的标记用  $Y_I$  来表示,如果  $Y_I = 1$ ,则包中至少有一个示例的标记为 1;如果  $Y_I = -1$ ,则表明包中所有示例的标记都为 -1。这可以表示为

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \quad \forall I, \text{ s.t. } Y_I = 1$$

$$y_i = -1, \quad \forall I, \text{ s.t. } Y_I = -1$$
(5-78)

在 MI-SVM 分类方法中，主要是从包的层面去考虑的，是通过最大化包的间隔来实现分类的，由于训练集中给定的是包中示例的特征向量，并没有直接描述包的特征，所以最大化包的间隔是通过最大化某些示例的间隔来代替的，具体方法是通过最大化正包中最有可能为正的示例和负包中最不可能为负的示例间的间隔来表示两类间的分类间隔。具体的表达式为

$$\min_{(\mathbf{w}, \mathbf{b}, \xi)} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_I \xi_I$$

$$\text{s.t. } \forall I, Y_I \max(\langle \mathbf{w} \mathbf{x}_i + \mathbf{b} \rangle) \geq 1 - \xi_I, \quad \xi_I \geq 0$$
(5-79)

由于负包中示例的标注是已知的，所以可以表示为当  $Y_I = -1$  时， $-(\langle \mathbf{w} \mathbf{x}_i + \mathbf{b} \rangle) \geq 1 - \xi_I$ ，对于正包示例，即  $Y_I = 1$ ，选择一个变量  $S(I) \in I$  作为正示例包  $B_I$  的描述，这时对正包示例的约束可以表示为  $(\langle \mathbf{w} \mathbf{x}_{S(I)} + \mathbf{b} \rangle) \geq 1 - \xi_I$ 。所以，可以将上述约束分为两部分来写，这样式 (5-79) 就可以表示为

$$\min_S \min_{(\mathbf{w}, \mathbf{b}, \xi)} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_I \xi_I$$

$$\text{s.t. } \forall I, Y_I = -1 \wedge -(\langle \mathbf{w} \mathbf{x}_i + \mathbf{b} \rangle) \geq 1 - \xi_I, \quad \forall i \in I$$

$$Y_I = 1 \wedge \langle \mathbf{w} \mathbf{x}_{S(I)} + \mathbf{b} \rangle \geq 1 - \xi_I, \quad \xi_I \geq 0$$
(5-80)

与 MI-SVM 不同，mi-SVM 是从示例层来考虑数据的优化问题的，它将标准的 SVM 转化为

$$\min_{y_i} \min_{(\mathbf{w}, \mathbf{b}, \xi)} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i$$

$$\text{s.t. } \forall i, y_i (\langle \mathbf{w} \mathbf{x}_i + \mathbf{b} \rangle) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, l$$
(5-81)

$$\sum_{i \in I} \frac{y_i + 1}{2} \geq 1, \quad \forall I, \text{ s.t. } Y_I = 1 \quad y_i = -1, \quad \forall I, \text{ s.t. } Y_I = -1$$

在 mi-SVM 中，由于正包示例的具体标注是不确定的，所以标注  $y_i$  也为未知量，通过求解式 (5-81)，可以同时获得最优的超平面和模式的确切标注。图 5.12 所示是分别用 mi-SVM 和 MI-SVM 分类的两个例子，其中数字为正包中的示例，“-”为负包中的示例。由图 5.12 中可以看出，mi-SVM 得到的分类超平面可以确保所有的负示例都能够被划分到负半平面，而正包中至少有一个示例被划分到正半平面。

在 mi-SVM 的迭代算法中，由于 SVM 核函数的强大功能，很多时候迭代仅被执行一次就结束了，所以 mi-SVM 只是得到了一个可行解，并不是全局最优解。针对这个问题，路晶等<sup>[92]</sup>提出了启发式 SVM 多示例学习 HSVM-MIL，该算法将表示图像语义内容的关键词当作图像类别标签，自动标注问题转化为图像分类问题。在 mi-SVM

算法的每次迭代中, HSVM-MIL 尝试改变其中一个样例的类别标号来防止 mi-SVM 过快跳出迭代, 并且要求改变标号后的迭代分类风险要小于上次迭代, 从而起到优化目标函数的作用。

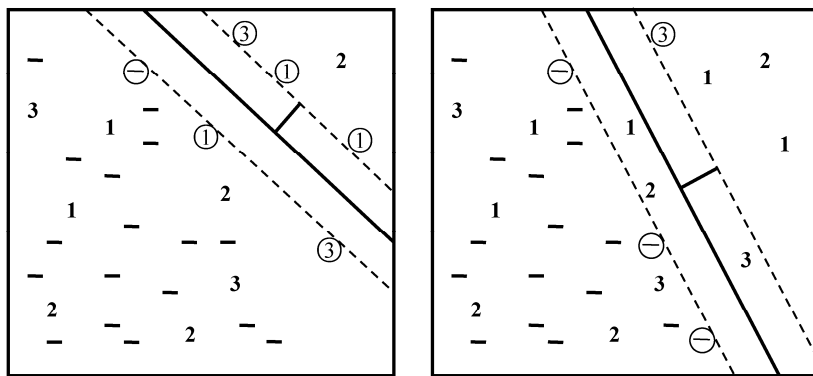


图 5.12 mi-SVM 和 MI-SVM 分类的描述

基于非对称支持向量机的多示例学习算法 (Asymmetrical Support Vector Machine-based MIL, ASVM-MIL)<sup>[93]</sup>也是通过扩展 SVM 来实现多示例图像标注的一种方法。ASVM-MIL 通过对错分的正例和负例引入不同的惩罚函数将 SVM 应用于多示例图像分类。在 ASVM-MIL 中, 为了将支持向量机能够应用于多示例分类问题, 需要最小化包的分类误差, 即

$$\begin{aligned} & \text{minimize } \langle \mathbf{w} \cdot \mathbf{w} \rangle + CE_B \\ & \text{s.t. } y_i(\mathbf{w}x_i + \mathbf{b}) \geq 1 - \xi_i, \xi_i \geq 0, y_i \in \{\pm 1\} \end{aligned} \quad (5-82)$$

其中,  $E_B$  表示错分的包的数量。为了求解式 (5-82), 需要将分类误差从包层具体到示例层, 然而由于示例的确切标注是未知的, 所以无法直接得到  $E_B$ 。这里,  $E_B$  代表两类错分误差,  $E^+$  表示错分为正示例的个数,  $E^-$  则表示错分为负示例的个数。ASVM-MIL 的目标是对错分为正示例和错分为负示例引入不同的损失函数。

$$\begin{aligned} & \text{minimize } \langle \mathbf{w} \cdot \mathbf{w} \rangle + C_1 E^+ + C_2 E^- \\ & \text{s.t. } y_i(\mathbf{w}x_i + \mathbf{b}) \geq 1 - \xi_i, \xi_i \geq 0, y_i \in \{\pm 1\} \end{aligned} \quad (5-83)$$

其中,  $C_1$  和  $C_2$  分别表示不同的惩罚因子, 由于正包中正示例的个数是未知的, 理论上来说是无法确定  $C_1$  和  $C_2$  的。但是, 可以假设正包中正示例的个数是大于 1 的, 所以, 当其中某个正示例被错标为负时, 不会导致包的标记错误。然而, 如果某个负示例被错标为正, 则会导致整个包的标记错误。通过分析, 可以得出  $C_1$  应该大于  $C_2$ , 不失一般性地, 假设  $C_2 = 0$ , 则有

$$\begin{aligned}
& \text{minimize } \langle \mathbf{w} \cdot \mathbf{w} \rangle + C \sum_{i=1}^l \xi_i^2 \\
& \text{s.t. } y_i(\mathbf{w}x_i + \mathbf{b}) \geq 1 - \frac{(y_i + 1)}{2} \xi_i, \xi_i \geq 0, y_i \in \{\pm 1\}
\end{aligned} \tag{5-84}$$

最后，仍然使用启发式迭代算法来求解式 (5-84)。

MI-SVM、mi-SVM、HSVM-MIL 和 ASVM-MIL 四种扩展的支持向量机算法中，后 3 种算法是从示例层面进行研究的，即可以实现图像区域的类别划分。HSVM-MIL 和 ASVM-MIL 是在 mi-SVM 的基础上提出的改进算法，所以这两种算法的图像分类效果要优于 mi-SVM。而 MI-SVM 是从包层面考虑的，无法确定图像中每个区域的标注，但是可以较好地得到整体图像的标注。

#### 4. 基于包层多示例学习的图像标注<sup>[94]</sup>

基于机器学习的图像标注方法研究的目的在于利用机器学习技术实现从图像的低层特征到语义概念间的映射，从而为两者之间的语义鸿沟搭建桥梁。其关键问题之一就是如何对图像的视觉内容赋予合适的描述。如果直接提取整幅图像的全局低层特征来表示图像的视觉内容，那么由于图像数据往往存在歧义性，将最终导致标注效果不好。图 5.13 所示是标注为“elephant”的 6 幅图像。从这些样图中可以明显地看出，图像中除了“elephant”对象以外，还有“sky”、“grass”等视觉区域，所以直接提取图像的全局特征无法准确、细致地表达图像蕴含的语义内容。针对上述问题，研究者将图像分割为若干个区域，用这些区域的特征来共同描述图像，这时，一幅图像就被表示为多个特征向量。由于对于已经过手工标注的图像数据，语义概念描述往往只是针对整幅图像的，而并不是针对图像中的具体区域的，也就是说，图像中的区域与描述图像的语义关键词不具备对应关系。因此，很多研究工作都把图像标注问题看成是多示例学习问题。



图 5.13 elephant 类图像

在多示例学习框架中，训练样本是由包含若干个示例的包构成的，包的标注是已

知的, 或者为正或者为负, 而包中示例的标注是未知的。如果一个包的标注为正, 则包中至少有一个示例的标注为正; 如果包的标注为负, 则表示包中示例的标注都为负。再来看自动图像标注问题, 将整幅图像看作包, 图像中的每个区域看作示例, 图像的标注是已知的, 图像中每个区域的标注是未知的, 所以自动图像标注问题是多示例学习问题。

为了解决多示例图像标注问题, 研究者提出了很多学习算法。一方面, 改造或扩展其他的机器学习算法用来解决多示例学习问题; 另一方面, 一些非常成熟而且有效的机器学习算法本身就具有很好的分类性能, 我们希望通过这些算法直接用于问题的处理。例如, 支持向量机, 它在解决小样本数据、高维模式识别和非线性问题中表现出很多特有的优势, 而且还可以推广应用到函数拟合等很多其他机器学习问题中。如果直接将支持向量机应用于图像标注的预测, 即将正包中所有示例都视为正示例, 负包中所有示例都视为负示例, 则会因为正包中噪声太多而严重影响预测性能。鉴于此, 这里主要研究如何将多示例学习问题转变为可以用现有机机器学习算法直接来进行处理的问题, 把由多个视觉特征向量表示的图像转变为单个特征向量描述, 图像由示例空间转化到包的空间, 在包空间实现图像的单一特征描述, 而包的标注是已知的, 这样就可以直接利用标注的支持向量机学习算法来实现新图像的预测。

在基于包层多示例学习的图像标注算法中, 首先需要构建包空间; 然后将多示例图像从示例空间映射到包空间, 在包空间每幅图像对应空间的一个点; 当图像集中每一幅图像都可以用包空间的一个特征向量来表示时, 就可以采用传统的支持向量机算法来实现自动图像标注了。

### 1) 包空间的构建

#### (1) 多样性密度算法构建包空间

Chen 等在文献[94]中采用多样性密度(DD)算法来学习构造包空间的“示例原型”, 算法将正包中出现而负包中不出现的示例称为是示例原型, 然后采用 DD 算法获取这些示例原型构建包空间。下面具体分析基于多样性密度算法构造包空间的方法。

已知一个数据集  $B = \{B_i^+, B_i^- \mid i=1, 2, \dots, m\}$ , 其中,  $B_i^+$  代表第  $i$  个正包,  $B_{ij}^+$  代表第  $i$  个正包的第  $j$  个示例,  $B_{ijk}^+$  代表第  $i$  个正包的第  $j$  个示例的第  $k$  个属性的值; 同理, 令  $B_i^-$ 、 $B_{ij}^-$ 、 $B_{ijk}^-$  分别表示第  $i$  个负包, 以及包中的第  $j$  个示例和该示例的第  $k$  个属性的值。Maron<sup>[78]</sup>提出的多样性密度算法在 5.3.3 小节中已有介绍, 该算法主要通过最大化式 (5-70) 的 DD 函数来寻找属性空间多样性密度最大的点。

通过求解式 (5-70), 可以得到最大多样性密度点。由于 DD 函数是连续的且是高度非线性的, 所以多样性密度空间中存在多个局部极小点, 普通的 DD 算法一般采用梯度上升法来寻找最大点。为了得到最大多样性密度点, 它们将每个正包示例都作为初始点进行搜索, 所以 DD 算法的运算开销很大。为了进一步分析, 将式 (5-70) 至

式 (5-73) 进行综合, 设  $d$  为多样性密度点, 则可得到

$$\begin{aligned} DD(x=d) &= \prod_i \Pr(d | B_i, l_i) = \prod_i [1 - |l_i - \text{Label}(B_i | d)|] \\ \text{Label}(B_i | d) &= \max_j \left\{ \exp \left[ - \sum_k w_k (B_{ijk} - d_k)^2 \right] \right\} \end{aligned} \quad (5-85)$$

从式 (5-85) 可以看出, DD 函数对负包示例非常敏感, 如果负包中某个示例接近多样性密度点  $d$ , 则  $\text{Label}(B_i | d)$  的值将接近 1, 这将导致  $DD(x=d)$  的值接近 0。也就是说, 如果负包中某个图像区域与正包中感兴趣点视觉效果比较相似的话, 则会导致多样性密度最大的点取得较小的数值, 从而漏掉某些 DD 值。而在实际图像集中, 表达不同语义的图像区域具有相似的视觉特征是难免的。例如, beach 类和 snow mountain 类图像, 图像中很多视觉区域相似度非常高, “water” 是 beach 类图像的典型代表区域, 它与 snow mountain 类中的 “mountain” 等区域的相似度比较高。在采用 DD 算法寻找示例原型时, 选择 beach 类图像为正包, 如果负包中包含 snow mountain 类图像, 则会大大削弱示例原型 “water” 的 DD 值, 所以最后的示例原型中可能不包括 “water” 这类的区域代表, 从而导致在构建包空间时会丢失一些很有代表性的视觉区域。另外, 由于 DD 算法为了得到最大多样性密度点, 通常选择正包中所有示例作为寻优起始点, 因此 DD 算法的训练时间开销相当大, 如果不这样做, 又难以找到理想的解。

通过上述分析总结可以发现, 由 DD 算法来构建包空间主要存在以下两个问题:

①当正包和负包中含有相似区域时, 可能会丢失一些示例原型; ②DD 算法在寻找多样性密度点时, 耗时多, 严重影响标注效率。针对 DD 算法构建包空间存在的缺点, 我们提出了基于视觉词汇构建包空间的方法。

## (2) 视觉词汇构建包空间

图像标注主要和两种不同的媒体相关, 一种是图像, 另一种是文本, 这两种媒体相互补充, 共同来表达图像本身要传递给人的信息。图像标注的目的就是建立这两种媒体映射关系, 即根据图像的视觉内容来确定相应的文本语义描述。描述图像语义的文本一般称为语义关键词或文本词汇, 词汇能够反映明确、清晰的语义信息, 是图像标注任务中相对固定的一种描述图像内容的方式。另一种媒体——图像, 一般采用图像的低层特征 (如颜色、纹理、形状等) 来描述。相比词汇, 图像的低层特征比较客观, 但是却无法表达图像所包含的清晰、明确的语义。人类对图像视觉内容的识别是基于图像中某个或几个感兴趣的目标区域的, 其中, 很多目标区域都能够表达一定的语义概念, 所以这里在低层视觉特征的基础上, 提取那些具有明确的语义概念所对应的图像区域来描述图像这种媒体。我们将这些区域定义为视觉词汇, 并采用这些视觉词汇来构建包空间。

之所以选择视觉词汇来构建包空间, 是因为每一类具有某个共同语义概念的图像都会包含一些典型的、特有的视觉词汇, 这些视觉词汇是支撑该语义概念的元素。我们可以把这些视觉词汇看作这类图像的属性。例如, 将图 5.14 中的图像进行分类, 类

别为 beach 和 elephant。我们很容易把图 5.14 (a) 中的 4 幅图像分为 beach 类，因为这几幅图像中有沙滩、海水、棕榈树，这些视觉词汇包含的语义概念是和 beach 类图像相关的；图 5.14 (b) 中的 4 幅图像被分为 elephant 类别，这是由于图像中的大象、草地、树木等视觉词汇可以和 elephant 类别联系到一起。

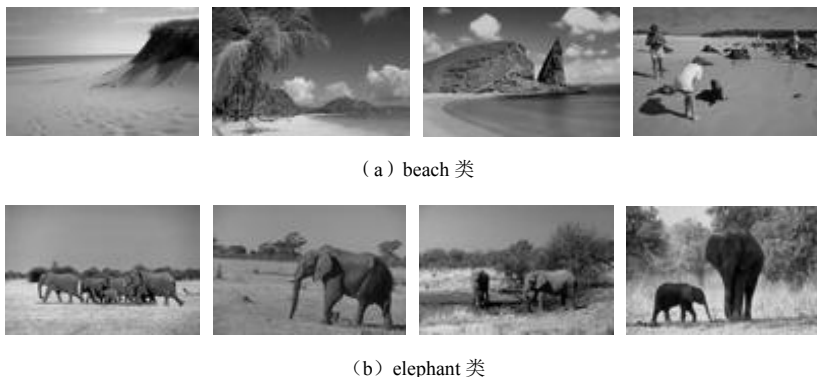


图 5.14 beach 类和 elephant 类图像示例

存在于人脑中的先验知识使我们很容易实现如上图像的分类，现在我们希望计算机系统能够像人脑一样，通过对已知标注图像的学习来实现未知图像的分类。首先，将所有包含高层语义概念的视觉词汇选择出来，然后和人脑认知过程类似，分析这些视觉词汇所能表达的语义概念。在图像集中，具有相同视觉内容的区域，其低层特征在示例属性空间都会聚集成一簇，视觉内容不同的图像区域，其低层特征在示例属性空间将分布在不同的簇中。如图 5.15 所示，分布比较密集的点表示这些示例（区域）是该类图像的典型示例（区域）。所以，我们用聚类算法对训练集中的示例进行聚类，将每个聚类视为一个视觉词汇。为了提取到更准确的视觉词汇，分别在各个图像类别中进行聚类，而并非将聚类算法直接应用于整个训练集。另外，为了得到有代表性、无冗余的视觉词汇，还需要考虑以下两个问题：①某类图像中，有些示例出现的频率很

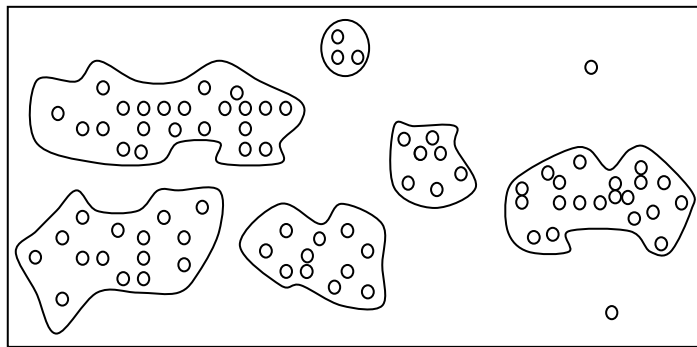


图 5.15 正包示例的分布

低，则将这些示例视为噪声去掉，不作为视觉词汇，如 beach 类中的“grass”、elephant

类中的“water”；②在正包和负包中会同时出现的示例，如图 5.14 中在 beach 类和 elephant 类中都经常出现的“sky”，说明它的类别区分度不高，也不选作典型的视觉词汇。

视觉词汇选择具体算法如下。

假设训练集  $B = \{(B_1, y_1), (B_2, y_2), \dots, (B_l, y_l)\}$  为已知标注的图像集。其中， $B_i$  表示已标注的图像，图像集中每幅图像被分割为若干个区域； $B_{ij}$  表示第  $i$  幅图像中的第  $j$  个示例（区域）。 $y_i$  表示图像的标注，取值为+1 或-1， $y_i = +1$  表示该图像属于某个语义概念类别； $y_i = -1$  则表示该图像不属于该语义概念类别。设  $S^+ = \{x_k^+ | k=1, 2, \dots, n\}$  表示训练集中所有标注  $y_i = +1$  的包中示例组成的集合， $S^- = \{x_k^- | k=1, 2, \dots, m\}$  则表示训练集中所有标注  $y_i = -1$  的包中示例组成的集合。具体算法步骤如下。

第 1 步：对集合  $S^+$  中的示例进行  $k$  均值聚类，先选择包含示例个数最多的包，然后以该包中示例的个数作为聚类个数  $k$  值，即

$$k = \max N_j \quad (5-86)$$

其中， $N_j$  表示包中示例的个数。

第 2 步：通过对正包示例聚类，得到  $k$  类别，用下式计算每个类别的类别中心。

$$Y_j = \frac{1}{N'_s} \sum_{i=1}^{N'_s} x_i \quad (5-87)$$

其中， $Y_j$  表示第  $j$  类的类别中心； $N'_s$  表示通过聚类后第  $j$  类所包含示例的个数； $x_i$  表示第  $j$  类的第  $i$  个示例。

第 3 步：聚类后，考察正包中每个类别包含的示例个数，如果某个聚类中包含的示例个数很少，即

$$N'_s \leq n_0 \quad (5-88)$$

则把这个类别作为噪声去掉。其中， $n_0$  为一个阈值。

第 4 步：计算保留下来的正包聚类中心  $Y_j$  分别与所有负包示例的距离，若

$$d = \|Y_j - x_i^-\| > d_0, j=1, 2, \dots, k', i=1, 2, \dots, m \quad (5-89)$$

则  $Y_j$  选择作为该类的典型视觉词汇。其中， $k'$  为去掉噪声后正包所包含的类别数。在这里设一个阈值  $d_0$ ，当正包示例聚类中心  $Y_j$  与负包示例  $x_i^-$  的距离小于阈值  $d_0$  时，认为该视觉词汇与负包示例  $x_i^-$  相似性较高，则它作为视觉词汇的区分度不高，所以不选择作为代表该类的典型视觉词汇。

通过上述方法提取的视觉词汇，通常是一簇具有相同视觉特征示例的中心点，对应一个明确的高层语义，因此这里将这些视觉词汇作为图像包空间的特征，用以构建一个新的空间——包空间。

## 2) 包空间投影特征计算

设  $V = \{v_1, v_2, \dots, v_C\}$  是由前面得到的  $C$  个区分度较强的典型视觉词汇，每个视觉



词汇都对应一定的高层图像语义概念，如果对每一幅图像计算它与所有视觉词汇间的相似度，这样一幅图像可以表示为一个  $C$  维特征向量，每一维都表示该图像与一个相应语义概念间的相似度。由于每幅图像都可以用一个  $C$  维特征向量来表示，而且训练集中图像的文本标注是已知的，所以可以用标准的 SVM 来进行学习。这里采用欧式距离来衡量图像与视觉词汇间的相似度。图像  $B_i = \{B_{ij} \mid j=1, 2, \dots, N_i\}$ ，图像与第  $k$  个视觉词汇间的相似度为<sup>[94]</sup>

$$d(v_k, B_i) = \max_{j=1, 2, \dots, N_i} \exp\left(-\|B_{ij} - v_k\|^2\right) \quad (5-90)$$

则图像  $B_i$  可以用下面的特征向量来表示。

$$\phi(B_i) = \begin{bmatrix} d(v_1, B_i) \\ d(v_2, B_i) \\ \vdots \\ d(v_C, B_i) \end{bmatrix} \quad (5-91)$$

其中， $\phi(B_i)$  为可以描述图像  $B_i$  的一个新的  $C$  维特征向量； $B_{ij}$  表示包  $B_i$  中的示例； $N_i$  表示包  $B_i$  中共有  $N_i$  个示例。式 (5-90) 计算的是图像包  $B_i$  中每个示例分别与视觉词汇  $v_k$  间的最小距离，也就是计算包中示例与每个视觉词汇间的相似度。

所有图像都被分割为若干个区域，每类图像都有自己典型的代表区域，通过聚类算法及一些约束条件选择出典型视觉词汇，可以得到具有很好语义概念的视觉词汇集  $V = \{v_1, v_2, \dots, v_C\}$ 。通过计算图像与视觉词汇间相似度来将多示例表达的图像转换成可由一个特征向量描述，即每幅图像只用一个特征向量来表示，这时多示例学习问题就转换成了单示例学习问题，可以用监督学习算法来实现图像的自动标注。

### 3) 图像包标注预测

支持向量机结构简单，具有全局最优性和较好的泛化能力，是求解模式识别和函数估计问题的有效工具，也是典型的监督学习算法之一。监督学习要求训练数据集中样本与标注的模式是明确的对应关系，而多示例学习问题中每个有标注的样本都由多个示例描述，所以无法直接将支持向量机用于多示例学习问题。通过前边介绍的转换，我们将多示例图像嵌入成包空间的一个点，变成了单示例学习问题，则可以用标准的 SVM 进行监督学习。

由前述的计算可以得到视觉词汇集  $V = \{v_1, v_2, \dots, v_C\}$ ， $C$  的取值不同，描述图像的单示例特征向量的维数就不一样，而不同维数的特征向量也往往具有不同的表达能力。为了充分利用这些信息，这里选择不同的参数  $n_0$ 、 $d_0$ ，就可以得到不同数量的视觉词汇，然后将多示例图像转换成不同长度的单示例描述，利用不同长度的单示例图像可以分别训练不同的 SVM 分类器，我们通过这些分类器集成来进行最终的图像类别判断。

假设图像的训练集为由  $l$  个样本对构成的数据集  $T$ ： $T = \{[\phi(B_1), y_1], [\phi(B_2), y_2], \dots,$

$[\phi(B_i), y_i]\}$ 。其中,  $\phi(B_i)$  是由前边推导得到的图像在包空间的特征向量;  $y_i \in \{+1, -1\}$  为相应的图像的标注。SVM 的目标是设计一个超平面, 将所有的训练样本正确分类。具有最大分类间隔的最优超平面可以使分类误差最小, 所以最优分类面的构造转换成求解下面的问题。

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{b}, \xi} & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} & y_i [\mathbf{w} \cdot \phi(B_i) + \mathbf{b}] \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned} \quad (5-92)$$

其中, 参数  $C$  为错误分类的惩罚因子;  $\xi_i$  为松弛变量;  $\phi(B_i)$  为图像  $B_i$  的特征向量。

采用拉格朗日乘子法求解这个具有线性约束的二次规划问题, 则上式转换成

$$\begin{aligned} \max & \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K[\phi(B_i), \phi(B_j)] - \sum_{i=1}^l \alpha_i \\ \text{s.t.} & \sum_{i=1}^l \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l \end{aligned} \quad (5-93)$$

其中,  $\alpha_i$  是拉格朗日乘子;  $K[\phi(B_i), \phi(B_j)] = \phi[\phi(B_i)] \phi[\phi(B_j)]$  是核函数。式 (5-93) 是求解不等式约束下的二次规划问题, 根据最优化理论中的 KKT 条件, 可得最优分类判别函数为

$$\text{label}(B) = \text{sign} \left\{ \sum_{i=1}^l \alpha_i y_i K[\phi(B_i), \phi(B_j)] + \mathbf{b} \right\} \quad (5-94)$$

当  $\text{label}(B)$  取值为 +1 时, 图像包  $B$  标注为正; 当  $\text{label}(B)$  取值为 -1 时, 图像包  $B$  标注为负。

由于采用了不同维数的单示例特征向量进行 SVM 训练, 所以可以得到多个 SVM 分类器, 最后将通过计算各个 SVM 分类器得到的后验概率的均值来判别图像类别。将式 (5-95) 右端均值最大的  $x_j$  作为图像类别, 即

$$\text{label}(B) = \max_{j=1,2,\dots,N} \arg \left( \frac{1}{m} \sum_{i=1}^m p_i(x_j | B) \right) \quad (5-95)$$

其中,  $p_i(x_j | B)$  为图像  $B$  分为类别  $x_j$  的概率;  $m$  为得到的分类器的个数;  $N$  为类别的个数。

综上所述, 本算法的具体步骤如下。

输入: 训练集为带有概念标记的图像集  $L$ , 测试集为待标注的图像集  $U$ 。

输出: 测试集  $U$  的标注。

初始化:  $V = \Phi$  (空集)。

第 1 步: 对  $L$  和  $U$  中的图像进行分割, 然后提取每个区域的低层特征。

第 2 步: 设定参数  $n_0$ 、 $d_0$ 。

第3步：选择视觉词汇， $M^+$  为训练集中属于某个语义概念的图像， $M^-$  为训练集中不属于该语义概念的图像， $M^+, M^- \subset L$ 。如果  $M^+$  中的视觉词汇  $v_k$  满足式 (5-88) 和式 (5-89)，则将  $v_k$  加入  $V$ 。

第4步：重复第3步，直到选出  $L$  中所有语义类别的视觉词汇  $V = \{v_1, v_2, \dots, v_C\}$ 。

第5步：用式 (5-90)、式 (5-91) 将  $L$  和  $U$  中的每幅图像都转换成  $C$  维的特征向量。

第6步：用  $L$  转换后的特征向量训练 SVM 分类器，同样，用  $U$  转换后的特征向量来测试 SVM 的分类性能。

第7步：清空  $V$ ，改变参数  $n_0$ 、 $d_0$ ，则将改变特征向量的长度  $C$ 。重复第3步至第5步，得到多个 SVM 分类器。

第8步：由式 (5-95) 确定未标注图像集  $U$  的标注。

## 5. 多示例多标记学习的自动图像标注

将多示例学习应用于自动图像标注可以有效地处理全局视觉特征表达的模糊性问题，然而，利用多示例学习来解决图像标注问题的方法往往只考虑单一语义概念的学习。真实世界的图像表达的内容是很难用单一语义关键词来描述的，自动图像标注本质上是个多语义概念问题。如图 5.16 所示，如果仅用“elephant”来描述图像，则会丢失图像中的其他语义信息，如“tiger”、“grass”等。当低层视觉特征相似程度较高时，不可避免地会出现对应的语义标注错误的现象，所以，自动图像标注技术必须考虑由视觉特征相似而导致的歧义性问题。在以往的研究中，多示例图像标注问题通常被分解为多个单语义学习问题，忽略了语义概念间本身所具有的相关性，而语义概念间的相互联系恰恰是解决由视觉信息造成标注语义歧义的有效方法。因而，利用语义概念间的相关特性无疑可以更好地提高图像标注性能，减少标注歧义的影响。



图 5.16 包含多个对象的图例

### 1) 多示例多标记学习的概述

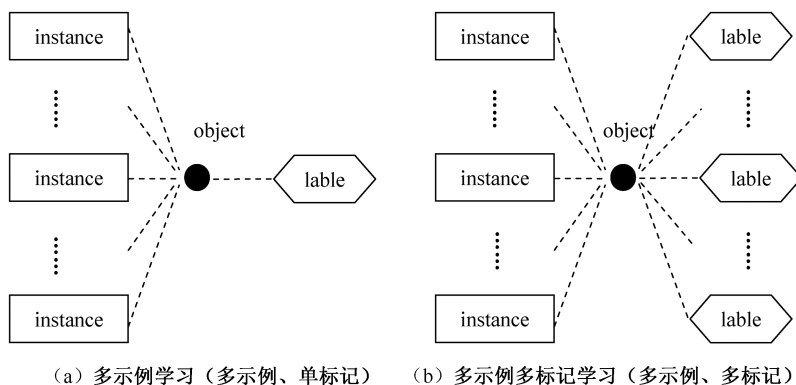
在利用机器学习技术解决图像标注问题时，比较常用的做法是先对图像区域或图像中的某个目标进行特征提取，用一个特征向量来描述这个区域，这样就得到了一个

示例,然后把示例与该对象所对应的类别标记关联起来,就得到了一个模式。在拥有了一个较大的例子集合之后,就可以利用某种学习算法来学得示例空间与标记空间之间的一个映射,该映射可以预测未标注示例的标记。假设每个对象只有一个类别标记,设 $\mathcal{X}$ 为示例空间, $\mathcal{Y}$ 为标记空间,则学习任务是从数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 中学得函数 $f: \mathcal{X} \rightarrow \mathcal{Y}$ ,其中 $x_i \in \mathcal{X}$ 为一个示例,而 $y_i \in \mathcal{Y}$ 为示例 $x_i$ 所对应的类别标记。在待学习对象具有明确的、单一的语义时,上面的学习框架已经取得了巨大的成功。然而,实际工作中常常涉及多个示例,以往的方法通常以图像整体为单元,通过提取图像的全局特征来描述图像的视觉内容,即用一个示例来表示图像。然而,一幅图像往往包含多个不同的视觉区域,这些视觉区域表述的是不同的语义概念,如图 5.16 所示,图像中包含了“elephant”、“tiger”、“grass”、“sky”等语义概念。这些语义是由图像中的不同区域表达的,如果仅仅提取全局特征,则不能细致准确地描述图像所包含的全部视觉内容。一幅图像包含若干个区域,每个区域都是一个示例,如果只用一个示例来描述整幅图像,则是一种简化,在表示阶段就丢失了很多有用的信息,这将为后续的学习阶段带来极大的困难。鉴于上述分析,多示例学习框架常常被用来描述自动图像标注问题。在多示例学习框架下,每幅图像被视为一个包,分割的图像区域被视为包中的示例,通过提取每个区域的低层特征来表达图像的视觉信息,利用多示例学习技术从正包和负包中建立自动标注模型。

前文从图像视觉特征分析发现图像低层特征适合选用多个示例来描述。从描述图像的语义关键词来看,图像中的内容一般需要用多个语义关键词描述,不适合用单一语义描述,因为单个语义关键词只能描述图像中的一个区域或一部分内容,并不能表述整幅图像的全部内容。仍以图 5.16 为例,图像中包含了“elephant”、“tiger”、“grass”、“sky”等语义内容,如果仅以其中单个关键词来描述图像内容,就无法充分描述出图像要表现的全部视觉信息。所以,图像标注本质上是一个多标记问题,也称作多语义概念学习问题。下面将给出多示例多标记学习的定义<sup>[95]</sup>。

在多示例语义标记的框架中, $\mathcal{X}$ 表示示例空间, $\mathcal{Y}$ 表示标记空间,学习的任务是由数据集 $\{(X_1 Y_1), (X_2 Y_2), \dots, (X_m, Y_m)\}$ 得到目标函数 $f_{\text{MIML}}: 2^{\mathcal{X}} \rightarrow 2^{\mathcal{Y}}$ 。其中, $X_i \subseteq \mathcal{X}$ 是示例集 $\{x_1^{(i)}, x_2^{(i)}, \dots, x_{n_i}^{(i)}\}$ , $x_j^{(i)} \in \mathcal{X}$  ( $j=1, 2, \dots, n_i$ ); $Y_i \subseteq \mathcal{Y}$ 是标注集 $\{y_1^{(i)}, y_2^{(i)}, \dots, y_{l_i}^{(i)}\}$ , $y_k^{(i)} \in \mathcal{Y}$  ( $k=1, 2, \dots, l_i$ )。这里 $n_i$ 表示 $X_i$ 中包含的示例个数; $l_i$ 则是 $Y_i$ 中标记的数量。为了更加清晰地理解多示例学习和多示例多标记学习,给出它们的学习框架,如图 5.17 所示。

为了解决多示例多标记学习问题,Zhou<sup>[96]</sup>等提出了 MIML-SVM 和 MIML-BOOST 两种算法。其中, MIML-SVM 以多标记学习为纽带,将多示例多标记学习问题转化为适合监督学习的单示例学习问题;而 MIML-BOOST 算法则是将多示例多标记学习转化为多个多示例学习问题。

图 5.17 两种机器学习框架<sup>[97]</sup>

Zhou 的这两种算法都没有考虑标注词间的相互关系。我们知道，图像的视觉数据往往具有歧义性，尤其是那些低层特征相似程度比较高的区域，由于目前多数标注方法在图像匹配时计算特征向量之间的欧氏距离，当图像视觉内容比较相近时，在特征空间，它们之间的距离会相对较小，这样的图像常常被错分为相同的类别，即使它们表达的是完全不同的语义内容。这是消弱图像标注精确度的一个很大的因素。例如，“ocean”和“sky”这两类图像区域，它们的低层特征在颜色上都是蓝色，相似程度很高，因此仅由低层特征是很难区分这两类图像的。

多语义图像标注的目标是得到图像的一组标注词汇，这些词汇之间并不是相互独立的，它们之间相互联系，而它们之间的相互联系恰是消除标注歧义的有效方法。例如，“grass”和“sky”共同标注一幅图像的概率远远大于“grass”与“ocean”；若可以确定图像中一个标注词是“grass”，那么图像的另一个标注词为“sky”的概率就远大于“ocean”。所以，利用标注词之间的关系来有效消除标注歧义问题是这里主要研究的内容，标注框架如图 5.18 所示。

从框架结构可以看出，自动图像标注分为如下两个过程：图像初始语义标注过程和标注改善过程。在图像初始标注阶段，将每个语义关键词视为一类，采用改进的多样性密度算法来学习语义关键词所对应的图像区域，然后结合贝叶斯分类模型为图像建立多语义描述。改进的多样性密度算法不仅可以改善图像的标注精确度，而且还可以有效地减少运算时间；在图像改善过程中，充分利用语义概念之间的相关性，对多个候选标注词进行修正，最终达到消除标注歧义、提高标注模型精度的目的。在图 5.18 中，输入到映射模型中的待标注样本是图像区域的特征。

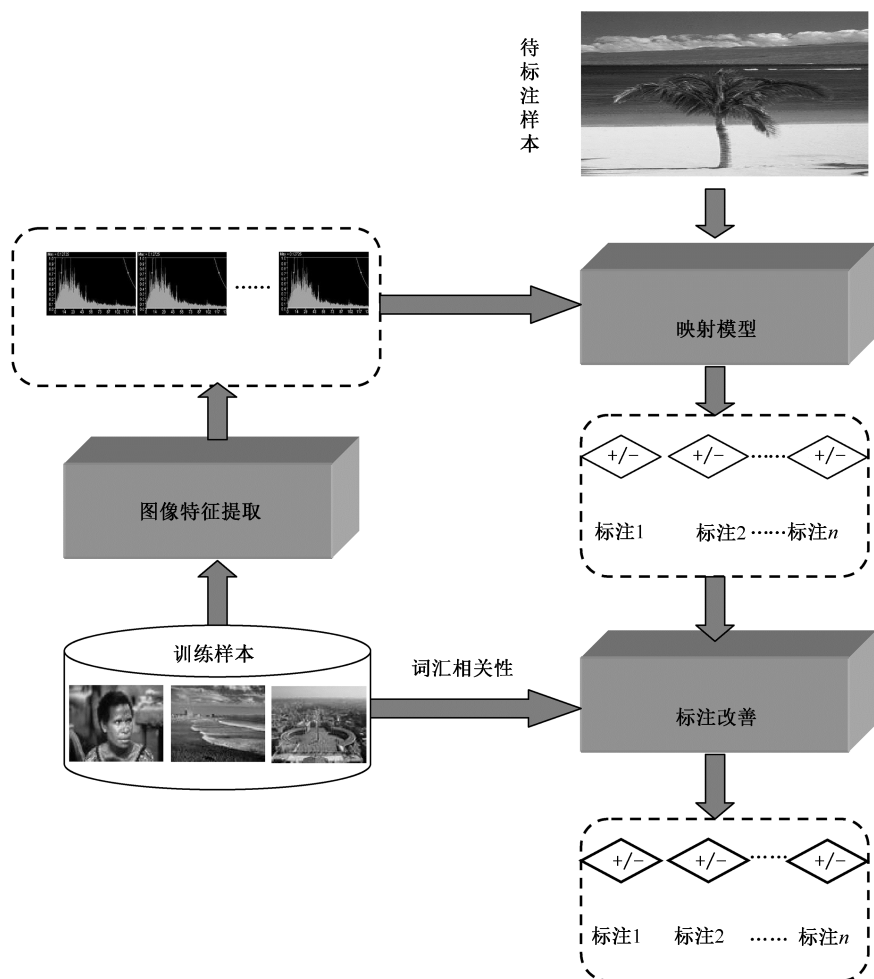


图 5.18 多示例多标记学习框架

## 2) 基于贝叶斯框架的多语义概念标记的学习

首先来介绍图像多语义概念标记的学习方法，即建立多示例图像与多语义概念之间的映射关系。这里采用改进的多样性密度算法来建立图像区域与语义关键词之间的对应关系，然后结合贝叶斯分类模型确定图像的多个标记。贝叶斯分类算法是统计学的一种分类方法，它是一类利用概率统计知识进行分类的算法。在许多场合，朴素贝叶斯 (Naive Bayes, NB) 分类算法可以与决策树和神经网络分类算法相媲美，该算法能运用到大型数据库中，而且方法简单，分类准确率高，速度快。贝叶斯分类模型将直观的知识表示形式与概率理论有机结合，是模式分类中进行不确定性推理和建模的有效工具。相比其他分类方法，贝叶斯模型具有以下一些优点。

(1) 将直观的知识表示形式与统计学理论知识有机结合，克服了某些模型不确定性推理的弱点。

(2) 可以方便地将先验知识和后验数据有机结合。

(3) 因为贝叶斯模型反映的是整个样本空间中的统计关系，所以它能够较好地处理不完备数据集，即缺少某一数据仍然可以建立精确的模型。

由于复杂贝叶斯模型的学习和推理均是一个 NP 问题，所以在处理实际问题时，一般采用结构简单、效率较高的朴素贝叶斯模型及其扩展模型作为分类器，以取得分类性能和分类效率间的折中。贝叶斯分类器的分类原理是通过某对象的先验概率，利用贝叶斯公式计算其后验概率，即该对象属于某一类的概率，选择具有最大后验概率的类作为该对象所属的类。目前，研究较多的贝叶斯分类器主要有 4 种：Naive Bayes（朴素贝叶斯）、TAN（树增强型朴素贝叶斯算法）、BAN 和 GBN。朴素贝叶斯分类模型采用最简单的网络结构，是一个简单、有效的分类器，它可以预测给定样本属于某个类别的概率。

朴素贝叶斯分类模型假设特征变量与给定类的影响独立于其他特征，即特征独立性假设。假设变量集  $Y = \{X_1, X_2, \dots, X_n, C\}$ ，所有的属性变量  $X_i$  ( $i=1, \dots, n$ ) 都条件独立于类变量  $C$ ，即每一个特征变量都以类变量作为唯一的父节点。图 5.19 比较直观地描述了朴素贝叶斯分类模型的结构特点。

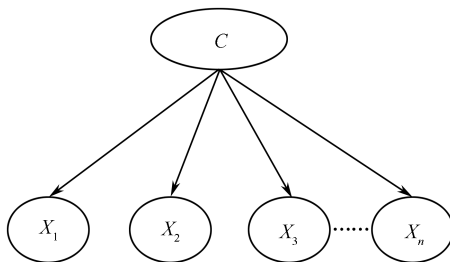


图 5.19 Naive Bayes 模型

假设训练集中每个样本可用一个  $n$  维特征向量来描述， $d = \{x_1, \dots, x_n, c_j\}$ ， $c_j \in \{c_1, \dots, c_m\}$ ， $c_j$  为样本类别标记。则如果满足

$$p(c_j | d) > p(c_i | d), \quad 1 \leq i \leq m, i \neq j \quad (5-96)$$

样本  $d$  的标注就为  $c_j$ 。根据贝叶斯公式，有

$$p(c_j | d) = \frac{p(c_j)p(d | c_j)}{p(d)} = \frac{p(c_j)p(d | c_j)}{\sum_{k=1}^m P(d | c_k)P(c_k)} \quad (5-97)$$

为了将误差最小化，贝叶斯分类器选择具有最大后验概率的类，即如果  $p(c_j | x) = \max_k P(c_k | x)$ ，则选择  $c_j$  为样本类别。

下面介绍贝叶斯框架下多示例图像的多语义概念标记学习算法。训练集中的每幅图像都被分割为若干个区域，图像被视为包，图像中的区域被视为示例，如果图像中某个区域的语义内容为  $w_i$ ，则该图像的标注词包括  $w_i$ 。把每个语义概念视为一个类别，则图像标注问题就可以转化为图像分类问题。贝叶斯分类器是一种典型的基于统计方法的分类器，是根据变量与变量间的因果关系进行建模的方法。

假设  $B$  表示已标注的图像训练集， $T$  表示未标注的图像测试集，其中，每幅图像都被分割为若干个区域。 $w = \{w_1, w_2, \dots, w_n\}$  为图像标注的关键词集合。设未标注图像  $I \in T$ ，通过计算条件概率  $p(w_i | I)$  的取值来确定图像  $I$  的语义标注。假设图像  $I$  被分割为  $l$  个区域， $I = \{r_1, r_2, \dots, r_j, \dots, r_l\}$ ，其中  $r_j$  表示图像  $I$  的第  $j$  个区域。以每个语义关键词作为一个类别，将图像标注问题转化为图像分类问题。在贝叶斯框架下，选择使  $p(w_i | I) = \max_{k \in \{1, 2, \dots, n\}} p(w_k | I)$  最大的  $w_i$  作为图像的标注。后验概率  $p(w_i | I)$  可以通过类条件概率  $p(I | w_i)$  和先验概率  $p(w_i)$  计算得到，即

$$p(w_i | I) \propto p(I | w_i) p(w_i) \quad (5-98)$$

假设图像  $I$  中的区域之间是相互独立的，则有

$$p(I | w_i) = \prod_{j=1}^l p(r_j | w_i) \quad (5-99)$$

由式 (5-98) 和式 (5-99) 可得

$$p(w_i | I) \propto \prod_{j=1}^l p(r_j | w_i) p(w_i) \quad (5-100)$$

文献[72]采用下式计算图像  $I$  的标注关键词。

$$\begin{aligned} w' &= \arg \max_{w_i \in w} \{p(w_i | I)\} \\ &= \arg \max_{w_i \in w} \{\max_{r_j \in I} p(r_j | w_i) p(w_i)\} \end{aligned} \quad (5-101)$$

其中， $p(w_i) = |B_i| / |B|$ ， $|B_i|$  表示图像训练集中标注为  $w_i$  的图像的数量， $|B|$  表示训练集中图像的总数目。式 (5-101) 在计算时先假定已知某个语义概念  $w_i$ ，计算在  $w_i$  给定时，选出图像  $I$  中标注为  $w_i$  的概率最大的区域；然后用同样的方法，分别计算语义标注集中所有的语义关键词与该图像各个区域的相关度；最后进行排序，选择前  $N$  个  $w_i$  作为该图像的标注。这里标注个数  $N$  是一个比较重要的参数， $N$  如果选择较大，则标注集中可能会出现与图像内容不相关的词汇； $N$  如果选择较小，则有可能无法描述出图像所包含的语义概念。

另外，式 (5-101) 得到的标注结果并没有考虑词汇之间的相关性。由于对图像相似程度的衡量都是基于欧氏距离的，如果图像区域之间距离较近，即使实际上它们所表达的语义完全不一样，也会导致这两个区域得到同样的标注的可能性很高。

通过上述分析，为了更好地实现图像标注，首先对式 (5-101) 进行修正，然后根据语义概念之间的相关性对由贝叶斯分类模型得到的多个语义进行分析，将候选词



中相关性较高的标注词赋予图像, 即

$$w' = \arg \max_{w_i \in W} \{p(r_1 | w_i)p(w_i)\} \cup \arg \max_{w_i \in W} \{p(r_2 | w_i)p(w_i)\} \cup \dots \cup \arg \max_{w_i \in W} \{p(r_k | w_i)p(w_i)\} \quad (5-102)$$

由式(5-102)可以得到图像每个区域最有可能的标注, 这里的区域是指分割得到的那些较大的区域。由此, 得到图像的多个语义概念标注。

为了得到图像标注, 需要计算式(5-102)。其中计算 $p(r_j | w_i)$ 的取值是关键。我们将语义概念为 $w_i$ 时图像区域为 $r_j$ 的条件概率定义为该区域与代表语义概念 $w_i$ 典型区域 $x_i$ 之间的距离, 即

$$\begin{aligned} P(r_j | w_i) &= \exp\left(-\|r_j - x_i\|^2\right) \\ &= \exp\left[-\sum_k a_{ik}^* (r_{jk} - x_{ik})^2\right] \end{aligned} \quad (5-103)$$

其中,  $x_i$ 是可以代表语义类别 $w_i$ 的典型特征区域;  $r_{jk}$ 和 $x_{ik}$ 分别是区域 $r_j$ 和 $x_i$ 的第 $k$ 维特征。

在上述的计算中, 需要求解语义概念 $w_i$ 所对应的典型区域 $x_i$ , 即建立视觉区域与对应语义关键字之间的关联。文献[45]首先采用聚类的方法将图像的区域聚集为多个类别, 称为**blob**, 然后根据训练集中图像区域与语义概念的联合概率来建立它们之间的关系。这类方法对聚类算法非常敏感, 图像标注的效果受到聚类结果的影响严重。文献[72]采用多样性密度算法来建立语义关键字与其对应区域的关系。多样性密度算法是解决多示例学习问题的一个典型框架, 在属性空间中若某点距离所有正包(用户感兴趣图像)中的一个示例(目标概念)很近, 却远离所有反包(不感兴趣图像)示例, 则该点的多样性密度较大, 那么该点为用户感兴趣概念所对应的区域的概率就较大。所以, 多样性密度算法是建立图像视觉区域与语义概念标注之间关联的一种更加有效的方法。

由于多样性密度函数是连续的, 而且是高度非线性的, 所以多样性密度空间中存在多个局部极小值点, 普通的多样性密度算法在计算过程中一般采用梯度下降法来寻找最大多样性密度点。由于属性空间存在多个局部极小值点, 因此在寻优过程中, 多样性密度算法每次选择正包中的一个示例作为起始点, 最后通过遍历所有正包中的示例来得到最大多样性密度点。如果正包数目较多, 并且正包中的示例个数也较多, 那么以正包示例作为起点来进行寻优就会带来较大的冗余计算, 耗费大量运算时间。针对上述问题, 提出了一种有效的改进算法用于建立语义概念与视觉区域之间的关联<sup>[98]</sup>。

在普通的多样性密度算法中, 计算式(5-70)是我们求解最大多样性密度点的关键。一般多样性密度算法通过遍历所有正包示例来得到目标概念点, 需要耗费大量运算时间。我们知道, 在某一点附近出现的正包越多, 负包示例越远, 该点为最大多样性密度点的可能就越大。也可以这样理解: 最大多样性密度点是距离正示例较近的点,

而且该点应该远离负包示例。所以，在利用梯度下降法进行搜索时，希望寻优的起始点为正示例点或离正示例较近的点，这样可以快速而且准确地找到多样性密度点。然而，正包中所有示例的标注是未知的，我们只能确定正包中肯定存在正示例，具体哪个示例的标注为正无法得到，为了得到全局的最大多样性密度点，普通的 DD 算法是以正包所有示例为起始点进行搜索，计算量较大。

为了能够快速、有效地得到全局多样性密度点，选择合适的起始点是非常必要的。在多示例图像的训练集中，虽然无法知道正包示例的确切标注，但是可以知道负包中的示例标注都为负，这是确定信息。如果正包中存在的某个示例与负包中示例相似性很高，可以认为该示例为负示例，所以不把它作为寻优的起始点，从而缩减寻优起始点个数。基于以上分析，这里提出了一种利用负示例的确定标注来指导寻优起始点的选择方法，即在 DD 算法中，如果正包中的某些示例与负包中的一些示例相似性很高，我们可以认为它们是同一类型的示例，不用于梯度寻优的起始点，从而减少寻优算法的计算量。

设  $B = \{B_i^+, B_i^- | i=1, 2, \dots, m\}$  为已知标注的数据包集合，其中  $B_i^+$  表示正包， $B_i^-$  表示负包。设  $S^+ = \{x_1^+, x_2^+, \dots, x_{c^+}^+\}$  为正包示例组成的集合， $S^- = \{x_1^-, x_2^-, \dots, x_{c^-}^-\}$  为所有负包示例组成的集合。具体算法步骤如下。

输入：正、负包示例集合  $S^+ = \{x_1^+, x_2^+, \dots, x_{c^+}^+\}$  和  $S^- = \{x_1^-, x_2^-, \dots, x_{c^-}^-\}$ 。

输出：寻优起始点集合  $S$ 。

初始化： $S = S^+$ 。

第 1 步：计算示例集  $S^+$  和  $S^-$  中示例之间的距离。

第 2 步：如果两个示例间距离满足

$$d = \|x_i^+ - x_j^-\| < d_0 \quad (5-104)$$

将  $x_i^+$  从  $S$  中删除。其中， $d_0$  为我们预先设定的一个阈值， $i=1, 2, \dots, c^+$ ， $j=1, 2, \dots, c^-$ 。

第 3 步：以第 2 步中得到的示例集合  $S$  作为梯度下降法寻优的起始点，进行 DD 算法。

由于上述算法可以有效减少正包中与负示例相似的示例，所以可以较大程度减少寻优起始点的个数，从而可以提高多样性密度算法寻找语义概念对应区域的效率。另外，由于正包中剩下的示例与负示例距离较远，以它们为起始点有利于快速准确地搜索到最大多样性密度点。

由多样性密度算法得到语义概念所对应的视觉特征，然后将该特征向量带入式 (5-102)、式 (5-103)，可以得到未标注图像的语义关键词。算法描述如下

第 1 步：示例集  $S^+$  为训练集中标注词为  $w_i$  所对应的正包中所有示例组成的集合， $S^-$  则为标注词不包含  $w_i$  的负包中所有示例的集合。

第 2 步：计算示例集  $S^+$  和  $S^-$  中示例之间的距离，如果两个示例间距离满足公式  $d = \|x_i^+ - x_j^-\| < d_0$  ( $i=1, 2, \dots, c^+$ ， $j=1, 2, \dots, c^-$ )，将  $x_i^+$  从  $S^+$  中删除。其中  $d_0$  为我们预先

设定的一个阈值。

第3步：以第2步中得到的示例集合  $S$  作为梯度下降法寻优的起始点，进行 DD 算法，得到标注关键词  $w_i$  对应的多样性密度最大点  $x^*$  和属性权重  $a_i^*$ 。

第4步：重复第1步至第3步，直到找到每个标注词所对应的视觉区域。

第5步：将由上述步骤得到的结果带入式 (5-102)、式 (5-103)，为未标注图像计算多语义概念标注结果。

由上述算法得到了多示例图像的多个语义概念标记，在这些标记中难免会存在噪声标注、错误标注，所以在下面将介绍利用语义概念间关系改善标注的方法。

### 3) 语义概念间相互联系对标注的改善

在初始标注词中，可能存在一些由于图像低层特征相似而造成的错误标注，为此可以根据标注关键词之间的相关特性来修正初始标注错误。

目前，多数自动图像标注技术采用机器学习算法来建立图像和关键字之间的语义关联，在学习过程中，往往是通过计算图像低层特征的欧氏距离，来衡量图像间的相似程度，该类方法在图像语义标注中取得了较好的标注效果。但这种基于特征空间欧氏距离的标注方法，是按照图像低层特征的相似程度进行归类的，具有相同视觉特征的区域将被划分为一类，即使区域的语义完全不同，也可能用相同的关键字对图像进行标注。例如，图像内容为“snow mountain”与“ocean”的两个图像区域，虽然它们的低层特征的相似程度很高，但所表达的语义概念是完全不同的，这时仅靠图像的低层特征是很难准确推知图像所表达的语义概念的。与图像低层特征相比，文本信息相对比较简洁，它们之间具有更加明确、清晰的语义关系。例如，在图像多语义标注中，如果已知其中一个标注词为“ship”，那么我们就可以推断图像标注为“ocean”的可能性要远远大于标注为“snow mountain”的可能性。所以，在低层特征的基础上再结合这些词汇之间的相互关系，可以有效修正标注的歧义性问题，从而提高自动图像标注的性能。

这里主要分析以下3种图像标注关键词改善的方法：基于语义概念间统计相关性的方法、基于词典的方法、基于随机游走的方法。

#### (1) 基于语义概念间统计相关性的方法

在训练集图像中，语义词汇之间存在自然的联系，如果两个词汇共同出现在一幅图像中频率较高，则这两个词汇之间的语义相关性较强，如“tiger”与“grass”，“beach”与“sea”。这是由于两个经常共同出现的词往往代表了两个语义概念或物体之间具有密切的联系，其中之一的出现也意味着另一概念或物体出现的可能性比较高，从而两个词同时标注一幅图像的概率也较高。在数据挖掘方法中，语义概念间的相互联系是十分重要的信息来源，这些相互关系可以非常有效地消除数据的歧义性，从而提高模型的性能。然而，如果只是简单地对词汇共现的频数进行统计，效果并不显著。文献[99]采用了用于文档分类中 TF-IDF 类似的方法来计算词汇之间的相互关系，即

$$K_{wc}(w_1, w_2) = K_c(w_1, w_2) \times \lg \left( \frac{N_T}{n_i} \right) \quad (5-105)$$

其中,  $K_c(w_1, w_2)$  表示关键字  $w_1$ 、 $w_2$  在训练集中共同作为一幅图像的标注而出现的次数;  $N_T$  为训练集的大小, 它表示图像训练集中所包含的全部图像的数量;  $n_i$  为标注词中包含关键字  $w_1$  的全部图像的数量。

也可以利用词汇之间的互信息来衡量它们的相关性, 假设  $k_i \in \{0, 1\}$  表示词汇  $i$  出现与否。如果词汇  $i$  不出现, 则  $k_i = 0$ ; 如果词汇  $i$  出现, 则  $k_i = 1$ 。于是可以采用下式来计算训练集中词汇  $i$  出现的概率。

$$p(k_i = 1) = \frac{\sum_{n=1}^N y_{n_i}}{N} \quad (5-106)$$

词汇  $i$  不出现的概率为  $p(k_i = 0) = 1 - p(k_i = 1)$ , 词汇  $i$  与词汇  $j$  共同在一幅图像中出现的概率为

$$p(k_i = 1, k_j = 1) = \frac{\sum_{n=1}^N (y_{n_i} \wedge y_{n_j})}{N} \quad (5-107)$$

同理, 词汇  $i$  与词汇  $j$  不同时出现在一幅图像中的概率和其中一个词汇出现而另一个没有出现的概率分别为

$$p(k_i = 0, k_j = 0) = \frac{\sum_{n=1}^N [(1 - y_{n_i}) \wedge (1 - y_{n_j})]}{N} \quad (5-108)$$

$$p(k_i = 1, k_j = 0) = \frac{\sum_{n=1}^N [y_{n_i} \wedge (1 - y_{n_j})]}{N} \quad (5-109)$$

$$p(k_i = 0, k_j = 1) = \frac{\sum_{n=1}^N [(1 - y_{n_i}) \wedge y_{n_j}]}{N} \quad (5-110)$$

采用互信息来计算词汇  $i$  与词汇  $j$  之间的相关性, 则

$$MI(i, j) = \sum_{k_i, k_j \in \{0, 1\}} p(k_i, k_j) \lg \left[ \frac{p(l_i, l_j)}{p(l_i)p(l_j)} \right] \quad (5-111)$$

根据文献[100],  $MI(i, j)$  越大, 词汇  $i$  与词汇  $j$  间的相关性越强。利用互信息来衡量词汇之间的相关性可以表达更多、更丰富的信息, 因为它不仅可以考虑词汇之间的共现概率, 而且还考虑它们共同不出现和不共同出现的情况。

## (2) 基于词典 WordNet 的方法

在建立语义关键词之间的相互联系时, 基于语义概念间共生关系的方法主要是依赖训练集中的词汇, 这些词汇传递的是某个范围或某个“局部”的信息。而在表达更为广泛的语义关键词之间的联系时, 我们需要寻找一个词汇量较大、包含丰富语义信

息的词典。在自然语言处理领域，由普林斯顿大学研究开发的 WordNet 作为一种结构化电子词典得到了广泛的应用，WordNet2.1 已经收录了多于 11 万的词汇。在该词典中，词汇被组织成一个同义词网络，每个同义词集合都代表一个基本的语义概念，并且这些集合之间也由各种关系连接。所有这些关系则可以用来衡量词汇之间的语义相关性。在自然语言处理领域，学者们提出了很多计算词汇间相关性度量的方法，如 Barnard 等在文献[101]中比较了多种计算词汇相关性的方法，这些方法各有优点，而且不适合单独进行词汇相关性的度量。文献[102]提出了多种利用 WordNet 改善标注的方法，在候选标注中采用 LIN、JNC 和 BNP 这 3 种语义相关性度量方法修正标注结果。

JNC 方法是典型有效的度量方法之一。在 JNC 方法中，首先需要已知一个已按照语义进行过良好标注的词库，利用这个词库估计语义概念  $c$  出现的概率，有

$$p(c) = \frac{\text{Freq}(c)}{N} \quad (5-112)$$

其中， $\text{Freq}(c)$  为语义概念  $c$  出现的次数； $N$  为所有语义概念出现的次数。语义概念  $c$  的信息容量（Information Content, IC）为

$$\text{IC}(c) = -\lg p(c) \quad (5-113)$$

两个语义概念之间的语义相关性用下式来计算。

$$K_{\text{wnc}}(c_1, c_2) = \frac{1}{\text{IC}(c_1) + \text{IC}(c_2) - 2 \cdot \text{IC}[\text{lcs}(c_1, c_2)]} \quad (5-114)$$

其中， $\text{lcs}(c_1, c_2)$  是概念  $c_1$  和  $c_2$  之间的最低公共父节点。

### （3）基于随机游走的方法

为了便于选择合适的、准确的标注关键词，标注算法需要为每个候选关键词赋一个置信度分值。很多已有的算法就是根据这个置信度分值为图像进行标注的，如文献[44]、[45]、[57]等，它们都是通过计算下式确定图像最终标注词的。

$$\text{score}(w_i) = p(w_i | I) \approx p(w_i | b_1, \dots, b_m) \quad (5-115)$$

其中， $I$  为待标注图像； $b_1, \dots, b_m$  为图像  $I$  的区域。为了更好地利用候选关键词的置信度分值及词汇集的信息，文献[103]将图像标注改善过程视为一个图排序问题，并用随机游走的算法来进行求解。

设每个候选关键词  $w_i$  为图  $G$  的一个顶点，所有的顶点用带权值的边连接，每条边上的权值表达了两个词汇之间的相关性。如果将单个词汇  $w_i$  作为搜索关键词， $\text{num}(w_i)$  表示用该词汇在网络搜索引擎上搜到的图像数量。如果将两个不同的词汇  $w_i$  和  $w_j$  作为联合搜索关键词，则用  $\text{num}(w_i, w_j)$  表示由这两个词汇在网络搜索引擎上搜到的图像数量。词汇  $w_i$  和  $w_j$  之间的相关性定义如下。

$$\text{sim}(w_i, w_j) = \begin{cases} \frac{\text{num}(w_i, w_j)}{\min\{\text{num}(w_i), \text{num}(w_j)\}}, & \text{num}(w_i, w_j) \neq 0 \\ 0, & \text{num}(w_i, w_j) = 0 \end{cases} \quad (5-116)$$

如果是非网络图像,  $\text{num}(w_i)$  表示训练集中已标注为  $w_i$  的图像个数,  $\text{num}(w_i, w_j)$  则表示训练集中标注为词汇  $w_i$  和  $w_j$  的图像个数。用矩阵  $S$  来表示所有词汇之间的相关性。在采用随机游走算法计算改善图像标注时, 在每一个顶点  $w_i$  有两种选择: 一是向前游走到点  $w_j$ , 权值为  $c \times \text{sim}(w_i, w_j)$ ; 另一种选择是随机跳转到任意一个点, 即

$$F = cAF + (1 - c)Y \quad (5-117)$$

其中,  $cAF$  表示按照词汇之间相关性进行随机游走的选择;  $Y$  为对应任意选择一个点的概率分布。

上面研究了多示例多标记学习的自动图像标注方法, 该方法主要包括两个过程: 建立初始标注过程和标注改善过程。在初始标注过程, 我们采用贝叶斯分类模型结合改进的多样性密度算法来实现多示例图像的多语义概念标注。多样性密度算法是多示例学习框架中的一种典型算法, 为了改善多样性密度算法存在较大冗余计算这一缺点, 采用负包示例的标注已知这一确定信息指导多样性密度算法的寻优过程, 然后结合贝叶斯分类模型为图像建立多语义描述。改进的多样性密度算法不仅可以改善图像的标注精确度, 而且还可以有效地减少运算时间。在图像标注改善过程中, 充分利用语义概念之间的相关性, 对多个候选标注词进行修正, 最终达到提高标注模型精度的目的。

## 参 考 文 献

- [1] 钟义信. 信息科学原理[M]. 北京: 北京邮电大学出版社, 2002.
- [2] 钟义信. “理解论”: 信息内容认知机理的假说[J]. 北京邮电大学学报, 2008, 31(3): 1-8.
- [3] 钟义信. 机器知行学原理[M]. 北京: 科学出版社, 2007.
- [4] Ying Liu, Dengsheng Zhang, Goujun Lu, et al. A survey of content-based image retrieval with high-level semantics[J]. Pattern Recognition, 2007, 40(1):262-282.
- [5] Julia Vogel, Bernt Schiele. Semantic Modeling of Natural Scenes for Content-based Image Retrieval[J]. International Journal of Computer Vision, 2007, 72(2):133-157.
- [6] Lew, Sebe, Djeraba, Jain. Content-based Multimedia Information Retrieval: State of the Art and Challenges[J]. ACM Transactions on Multimedia Computing, Communications and Application, 2006:1-19.
- [7] Li X, Chen L, Zhang L, et al. Image annotation by large-scale content-based image retrieval[C] // The 14th Annual ACM International Conference on Multimedia, 2006:607-610.
- [8] 英济民. 基于互联网数据集的图像标注技术研究[D]. 合肥: 中国科学技术大学, 2009.

- [9] Thomas Lehmann, Thomas Deselaers, Henning Shubert, et al. Irma-A Content-based Approach to Image Retrieval in Medical Applications[C] // Information Resources Management Association International Conference, 2006:911-912.
- [10] Tommasi T, Orabona F, Caputo B. Discriminative cue integration for medical image annotation[J]. Pattern Recognition Letters, 2008:1996-2002.
- [11] Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009: 210-227.
- [12] Fang Y, Geman D, Boujemaa N. An Interactive System for Mental Face Retrieval[C] // The 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, 2005:193-200.
- [13] Jan Schietse, John P Eakins, Remco C Veltkamp. Practice and Challenges in Trademark Image Retrieval[C] // The 6th ACM International Conference on Image and Video Retrieval, 2007:518-524.
- [14] Das S. Fiber, wrappers and a boosting-based hybrid for feature selection[C] // The Eighteenth International Conference on Machine Learning, 2001:74-81.
- [15] Kohavi R, John G. Wrappers for feature subset selection[C] // Artificial Intelligence 97, 1997:273-324.
- [16] Yu L, Liu H. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution[C] // Proceedings of ICML, 2003:856-863.
- [17] 王博. 文本分类中特征选择技术的研究[D]. 长沙: 国防科学技术大学, 2009.
- [18] Yu L, Liu H. Efficient Feature Selection via Analysis of Relevance and Redundancy[J]. Journal of Machine Learning Research, 2004:1205-1224.
- [19] Swiniarski R W, Skowron A. Rough set methods in feature selection and recognition[J]. Pattern Recognition Letters, 2003:833-849.
- [20] Last M, Kandel A, Maimon O. Information-theoretic algorithm for feature selection[J]. Pattern Recognition Letters, 2001:799-811.
- [21] 武建华, 宋擒豹. 基于关联规则的特征选择算法[J]. 模式识别与人工智能, 2009, 22(2): 256-262.
- [22] Dash M, Liu H, Yao J. Dimensionality Reduction of Unsupervised Data[C] // Proceedings of ICTAI, 1997:532-539.
- [23] Mitra P, Murthy C A, Pal S K. Unsupervised Feature Selection Using Feature Similarity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002:301-312.
- [24] Covoes T F, Hruschka E R, Castro L N D, et al. A Cluster-Based Feature Selection Approach[C] // Proceedings of HAIS, 2009:169-176.

- [25] Zeng H, Cheung Y. A new feature selection method for Gaussian mixture clustering[J]. Pattern Recognition, 2009:243-250.
- [26] Zhang D, Zhou Z, Chen S. Semi-Supervised Dimensionality Reduction[J] // Proceedings of SDM, 2007:629-634.
- [27] Lei Wang. Image Annotation and Classification[D]. Dallas:The University of Texas at Dallas,2006.
- [28] Ran Li, Jianjiang Lu, Yafei Zhang, et al. Dynamic Adaboost learning with feature selection based on parallel genetic algorithm for image annotation[J]. Knowledge-Based Systems, 2010, 23(3):195-201.
- [29] Setia L, Burkhardt H. Feature selection for automatic image annotation[J]. Lecture Notes in Computer Science, 2006:294-303.
- [30] 汪廷华, 田盛丰, 黄厚宽. 特征加权支持向量机[J]. 电子与信息学报, 2009, 31(3): 514-518.
- [31] Wang L, Khan L. Automatic image annotation and retrieval using weighted feature selection[J]. Springer Science + Business Media, 2006, 29(1):55-71.
- [32] 王娜, 李霞. 基于类加权的双v支持向量机[J]. 电子与信息学报, 2007, 29(4): 859-862.
- [33] Wang X, Wang Y, Wang L. Improving fuzzy c-means clustering based on feature-weight learning[J]. Pattern Recognition Letters, 2004:1123-1132.
- [34] Yeung D S, Wang X. Improving Performance of Similarity-Based Clustering by Feature Weight Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002:556-561.
- [35] 李洁, 高新波, 焦李成. 基于特征加权的模糊聚类新算法[J]. 电子学报, 2006, 34(1):89-92.
- [36] 王科平, 王小捷, 钟义信. 加权特征自动图像标注方法[J]. 北京邮电大学学报, 2011, 34(5):34-37.
- [37] Kira K, Rendell L A. The Feature Selection Problem: Traditional Methods and a New Algorithm[C] // Proceedings of AAAI, 1992:129-134.
- [38] Kononenko I. Estimating Attributes: Analysis and Extensions of RELIEF[C] // Proceedings of ECML, 1994:171-182.
- [39] Wang Li J, J Z. Automantic Linguistic Indexing of Pictures by a Statistical Modeling Approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003,25(19):1075-1088.
- [40] Town C, Sinclair D. Content-based image retrieval using semantic visual category[J]. Cambridge: AT& T Laboratories, 2001.
- [41] Lin C, Wang S. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):464-471.



- [42] 张翔, 肖小玲, 徐光祐. 基于样本之间紧密度的模糊支持向量机方法[J]. 软件学报, 2006, 17(5):951-968.
- [43] 范昕炜, 杜树新, 吴铁军. 可补偿类别差异的加权支持向量机算法[J]. 中国图像图形学报, 2003, 18(9):1037-1042.
- [44] Duygulu P, Barnard K, de Freitas JFG, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary[C] // The European Conference on Computer Vision, 2002:97-112.
- [45] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models[J]. The Int'l ACM SIGIR, 2003:119-126.
- [46] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures[J]. Neural Information Processing Systems (NIPS), 2004:553-560.
- [47] Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2004:1002-1009.
- [48] Deerwester S C, Dumais S T, Landauer T K, et al. Indexing by Latent Semantic Analysis[J]. JASIS, 1990:391-407.
- [49] Hofmann T. Probabilistic Latent Semantic Indexing[C] // Proceedings of SIGIR, 1999:50-57.
- [50] Jeon J, Manmatha R. Using Maximum Entropy for Automatic Image Annotation[C] // Proceedings of CIVR, 2004:24-32.
- [51] Shi R, Chua T, Lee C, et al. Bayesian Learning of Hierarchical Multinomial Mixture Models of Concepts for Automatic Image Annotation[C] // Proceedings of CIVR, 2006:102-112.
- [52] Kang F, Jin R, Chai J Y. Regularizing translation models for better automatic image annotation[C] // Proceedings of CIKM, 2004:350-359.
- [53] Kang F, Jin R. Symmetric Statistical Translation Models for Automatic Image Annotation[C] // Proceedings of SDM, 2005.
- [54] 王斌. 图像检索中自动标注与快速相似搜索技术研究[D]. 合肥: 中国科学技术大学, 2007.
- [55] Monay F, Gatica-Perez D. On image auto-annotation with latent space models[C] // Proceedings of ACM Multimedia, 2003:275-278.
- [56] Monay F, Gatica-Perez D. PLSA-based image auto-annotation: constraining the latent space[C] // Proceedings of ACM Multimedia, 2004:348-351.
- [57] Blei D M, Jordan M I. Modeling annotated data[C] // Proceedings of SIGIR, 2003:127-134.

- [58] Blei D M , Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003:993-1022.
- [59] 夏利民, 谭立球, 钟洪. 基于信息瓶颈算法的图像语义标注[J]. 模式识别与人工智能, 2008, 21(6):812-818.
- [60] 王梅, 周向东, 张军旗, 等. 基于扩展生成语言模型的图像自动标注方法[J]. 软件学报, 2008, 19(9):29-37.
- [61] Li J, Wang J Z. Automantic Linguistic Indexing of Pictures by a Statistical Modeling Approach[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003,25(19):1075-1088.
- [62] Town C, Sinclair D. Content-based image retrieval using semantic visual category[J]. Cambridge: AT& T Laboratories, 2001.
- [63] Cusano Claudio, Ciocca Gianluigi, Schettini Raimondo. Image annotation using SVM[C] // Proceedings of SPIE, 2003:330-338.
- [64] 高隽, 谢昭. 图像理解理论与方法[M]. 北京: 科学出版社, 2009.
- [65] Vapnik V N. The Nature of Statiscal Learning Theory[M]. New York: Springer-Verlag, 1995.
- [66] Christopher M Bishop. Pattern Recongnition and Machine Leaning[M]. New York: Springer, 2006.
- [67] Hoi S C H, Jin R, Zhu J, et al. Semisupervised SVM batch mode active learning with applications to image retrieval[J]. ACM Trans. Inf. Syst., 2009, 21(1):1233-1248.
- [68] Guo B, Gunn S R, Damper R I, et al. Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification[J]. IEEE Transactions on Image Processing, 2008, 17(1):622-629.
- [69] Boiman O, Shechtman E, Irani M. In defense of Nearest-Neighbor based image classification[C] // Proceedings of CVPR, 2008.
- [70] Maji S, Berg A C, Malik J. Classification using intersection kernel support vector machines[C] // Proceedings of CVPR, 2008.
- [71] Vedaldi A, Gulshan V, Varma M, et al. Multiple kernels for object detection[C] // Proceedings of ICCV, 2009:606-613.
- [72] 胡洋. 最大间隔方法及其在图像检索中的应用[D]. 合肥: 中国科学技术大学, 2009.
- [73] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the Multiple Instance Problem with Axis-Parallel Rectangles[J]. Artif. Intell. 1997:31-71.
- [74] Yamakawa H, Maruhashi K, Nakao Y. Predicting Types of Protein-Protein Interactions Using a Multiple-Instance Learning Model[C] // Proceedings of JSAI, 2006:42-53.

- [75] Zafra A, Romero C, Ventura S. Predicting Academic Achievement Using Multiple Instance Genetic Programming[C] // Proceedings of ISDA, 2009:1120-1125.
- [76] Reiji T, Hisashi K. Prediction of protein-ligand binding affinities using multiple instance learning[J]. Journal of Molecular Graphics and Modelling, 2010:492-497.
- [77] Zhou Xuefeng, Ruan Jianhua, Zhang Weixiong. Promoter prediction based on a multiple instance learning scheme[C] // ACM Int. Conf. Bioinformatics Comput. Biol., 2010: 295-301.
- [78] Maron O, Lozano-Pérez T. A Framework for Multiple-Instance Learning[J]. Neural Information Processing Systems(NIPS), 1997:570-576.
- [79] Qi Zhang, Sally A Goldman, Wei Yu, et al. Content-based image retrieval using multiple-instance learning[C] // Proceedings of 19th Int'l Conference on Machine Learning, 2002:682-689.
- [80] Zhang D, Wang F, Shi Z, Zhang C. Interactive localized content based image retrieval with multiple-instance active learning[J]. Pattern Recognition, 2010:478-484.
- [81] Chiang J Y, Cheng S, Huang Y. Multiple-instance image database retrieval by spatial similarity based on Interval Neighbor Group[C] // Proceedings of CIVR, 2010:135-142.
- [82] Ueda N, K Saito. Parametric Mixture Models for Multi-Labeled Text[J]. Neural Information Processing Systems(NIPS), 2002:721-728.
- [83] H Kazawa, T Izumitani, H Taira, et al. Maximal Margin Labeling for Multi-Topic Text Categorization[J]. Neural Information Processing Systems (NIPS), 2004.
- [84] M L Zhang, Z H Zhou. Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization[J]. IEEE Transactions on Knowl. Data Eng., 2006:1338-1351.
- [85] Zhang Q, Goldman S A. EM-DD: An Improved Multiple-Instance Learning Technique[J]. Neural Information Processing Systems (NIPS), 2001:1073-1080.
- [86] Andrews S, Tsochantaridis I, Hofmann T. Support Vector Machines for Multiple-Instance Learning[C] // Proceedings of NIPS, 2002:561-568.
- [87] Dooly D R, Zhang Q, Goldman S A. Multiple-Instance Learning of Real-Valued Data[J]. Journal of Machine Learning Research, 2002:651-678.
- [88] Maron O, Ratan A L. Multiple-Instance Learning for Natural Scene Classification[C] // Proceedings of ICML, 1998:341-349.
- [89] Yang C, Lozano-Pérez T. Image Database Retrieval with Multiple-Instance Learning Techniques[C] // Proceedings of ICDE, 2000:233-243.
- [90] J R Foulds, E Frank. Speeding Up and Boosting Diverse Density Learning[C] // Proceedings of Discovery Science, 2010:102-116.

- [91] Foulds J R, Frank E. Revisiting Multiple-Instance Learning via Embedded Instance Selection[C] // Proceedings of Australasian Conference on Artificial Intelligence, 2008:300-310.
- [92] 路晶, 马少平. 使用基于多例学习的启发式 SVM 算法的图像自动标注[J]. 计算机研究与发展, 2009, 46(5):864-871.
- [93] Yang C, Dong M, Hua J. Region-based Image Annotation using Asymmetrical Support Vector Machine-based Multiple-Instance Learning[C] // Proceedings of CVPR, 2006, 2(1):2057-2063.
- [94] 王科平, 杨艺, 王新良. 包空间多示例图像自动分类[J]. 中国图像图形学报, 2013, 18(9):1093-1100.
- [95] Chen Y, Wang J Z. Image Categorization by Learning and Reasoning with Regions[J]. Journal of Machine Learning Research, 2004:913-939.
- [96] Zhou Z, Zhang M. Multi-Instance Multi-Label Learning with Application to Scene Classification[J]. Neural Information Processing Systems (NIPS), 2006:1609-1616.
- [97] Zhou Z, Zhang M, Huang S, et al. MIML: A Framework for Learning with Ambiguous Objects[C] // Proceedings of CoRR, 2008.
- [98] Wang Keping, Wang Xiaojie. Automatic Image Annotation Based on the Multiple-instance Learning[J]. Journal of Information and Computational Science, 2011, 13(7): 2781-2788.
- [99] 卢汉清, 刘静. 基于图学习的自动图像标注[J]. 计算机学报, 2008, 31(9):1629-1639.
- [100] Jawaharlal Karmeshu. Entropy Measures, Maximum Entropy Principle and Emerging Applications[M]. New York:Springer-Verlag, 2003.
- [101] Barnard K, Duygulu P, Forsyth D A, et al. Matching Words and Pictures[J]. Journal of Machine Learning Research, 2003:1107-1135.
- [102] Jin R, Chai J Y, Si L. Effective automatic image annotation via a coherent language model and active learning[C] // Proceedings of ACM Multimedia, 2004:892-899.
- [103] Wang C, Jing F, Zhang L, et al. Image annotation refinement using random walk with restarts[C] // Proceedings of ACM Multimedia, 2006:647-650.

## 子空间特征提取技术

在图像检索领域，经常遭遇高维数据，如果不进行有效的降维，就有可能产生所谓的维数灾难。基于子空间分析的特征提取方法是降维的有效手段之一。本章在对子空间特征提取方法进行介绍的基础上，引入了 3 种子空间特征提取方法。

### 6.1 概 述

#### 6.1.1 降维原因

在图像检索领域，经常遭遇高维数据：一方面，用于表征数据的维数越高，其包含的信息越丰富细致；另一方面，高维数据占用大量存储空间，计算量大，对存储与计算是个巨大的挑战。如何降低计算的负荷从而提供一个更加智能的模型得到了研究人员的广泛关注。对高维数据进行维数约减从而得到其简洁数据表示的过程称为降维。从广义上讲，降维的过程就是将高维数据投影到一个低维空间，同时保持高维数据内在的本质特征；从狭义上讲，降维的原因分为 3 类，即缓解维数灾难、建立有效的数据模型、降低识别的时间消耗<sup>[1]</sup>。

##### 1. 缓解维数灾难

降维的第一个动机是缓解维数灾难。“维数灾难”一词最早由 Richard Bellman 提出，是指在统计估计过程中，为达到相同的估计精度，所需的样本数随维数的增加而呈指数增长<sup>[2]</sup>。下面通过一个例子加以说明。假设  $X$  为一个包含 10 个样本的数据集，

数据  $X$  在一维、二维、三维空间内的描述如图 6.1 所示。

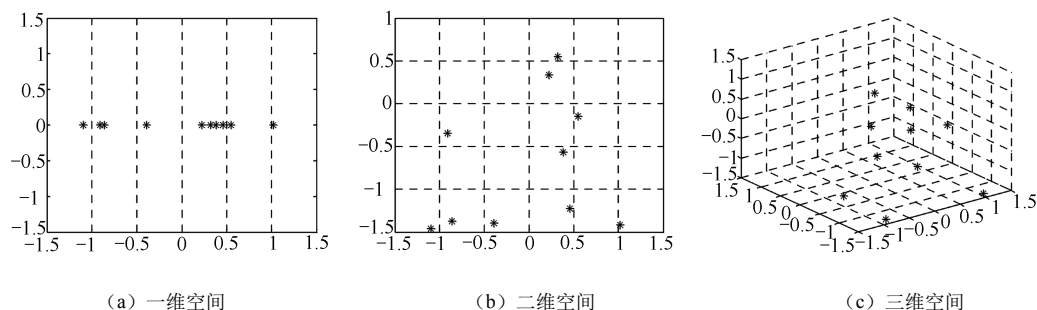


图 6.1 数据在不同维数空间内的描述

根据数据的空间描述，可以提取出数据的一些相关统计特性，这里采用直方图来表示数据的分布特性，即将数据空间分成若干个大小相同的区间，然后统计各个区间内样本点的个数。对于图 6.1 中的数据，将每一维空间分成大小相同的 6 个区间，那么一维空间、二维空间、三维空间分别被分成了 6 个、36 个、216 个区间。从图 6.1 中可以看出，在一维空间内，每一个区间至少包含一个样本，平均每个区间包含 1.66 个样本。在二维空间内，16 个区间没有样本，平均每个区间包含 0.55 个样本，这意味着若利用直方图对数据分布特性进行估计，在二维空间内需要的样本数是一维空间的 3 倍。同样，在三维空间内，情况更加糟糕，在 216 个区间内，绝大多数不包含任何样本，平均每个区间仅包含 0.09 个样本，这就意味着在三维空间内如果达到一维空间内对数据分布特性估计的精度，需要的样本数为一维空间的 18 倍。

因此，在统计估计的过程中重要的不是样本的个数，而是样本的密度。对于给定的样本规模，随着维数的不断增加其密度急剧下降，为了保持相同的密度，那么需要的样本的规模随着维数的增加呈指数增长。这时可以通过降维来保持数据的关键特性从而缓解维数灾难，一旦数据映射到低维子空间，我们就可以从同样规模的数据中提取出更加重要的信息。

## 2. 建立有效的数据模型

降维的第二个动机是建立有效的数据模型，加深对数据的理解。假设有一组关于手的图像<sup>[3]</sup>，如图 6.2 所示。

对于图 6.2 中的图像，若其分辨率为  $m \times n$ ，那么就需要在  $m \times n$  维空间内描述每一幅图像。但从图 6.2 中可以看出，该组图像的区别仅仅在于手指的弯曲程度变化及手腕的旋转程度变化，因此可以在二维空间内建立该组图像的模型，即可以利用二维空间中的一点表示一幅图像，其横坐标表示手腕的旋转程度，纵坐标表示手指的弯曲程度，这样就可以在二维空间内分析高维图像数据，一方面大大节省了存储空间，另一方面大大加深了对图像本质特征的理解。

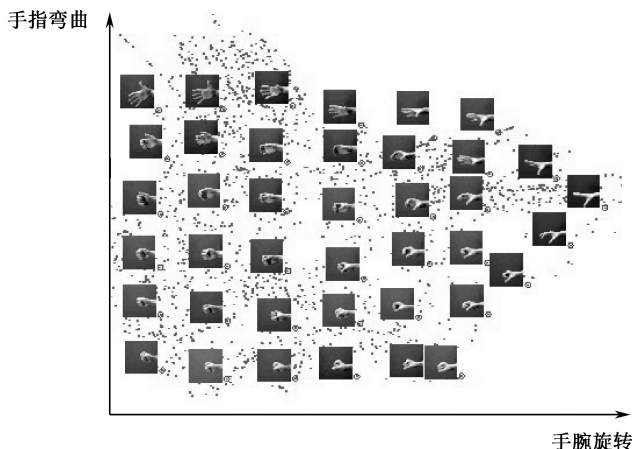


图 6.2 一组关于手的图像

### 3. 降低识别的时间消耗

降维的第三个动机是降低识别的时间消耗。模式识别的过程实际上就是待识别的对象与已知对象的匹配的过程，计算二者之间的相似度是最简单的一种匹配方法。若每个对象用一个一维向量进行表示，那么匹配的过程就演变为两个向量的点积运算。有实验表明，待识别对象的匹配过程的时间消耗随维数的增加也呈指数增长，因此有必要通过降维来降低对象识别的时间消耗。

## 6.1.2 子空间特征提取方法的形式化描述及分类

基于子空间分析的特征提取方法是降维的有效手段之一，其基本原理是通过线性或非线性映射将高维数据映射到低维空间，从而找出隐藏在高维数据中有意义的低维特征，在降维的过程中，尽量去除冗余信息，保留有意义的信息，使得降维后的信息损失最小。子空间特征提取方法的形式化描述如下<sup>[4]</sup>。

假设有样本集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i \in \mathbf{R}^m$ , 其中,  $N$  为样本的个数,  $m$  为样本的维数。样本  $\mathbf{x}_i$  的类别标号  $l_i \in L = \{1, 2, \dots, M\}$ , 其中,  $M$  为样本类别个数,  $n_c$  为属于第  $c$  类的样本个数。那么子空间特征提取方法就是寻找一个映射函数  $F$ , 使得  $\mathbf{y}_i = F(\mathbf{x}_i)$ , 其中,  $\mathbf{y}_i \in \mathbf{R}^{m'}$  为  $\mathbf{x}_i$  的低维映射,  $m'$  为低维映射  $\mathbf{y}_i$  的维数且满足  $m' < m$ 。

可以从不同的角度对子空间特征提取方法进行分类。

(1) 根据映射函数  $F$  是否为线性映射, 可以分为线性子空间特征提取方法和非线性子空间特征提取方法, 非线性子空间特征提取方法又可以分为核方法和流形方法。

(2) 根据在子空间特征提取的过程中是否利用先验知识(如类别标注信息、边信息、反馈信息), 可以分为有监督方法、无监督方法和半监督方法。

(3) 根据样本的表示形式, 可以分为向量方法和张量方法。

## 6.2 经典的子空间特征提取方法

### 6.2.1 线性方法

主元成分分析 (Principal Component Analysis, PCA)<sup>[5]</sup> 和线性鉴别分析 (Linear Discriminant Analysis, LDA)<sup>[6]</sup> 是两种最为经典的线性子空间特征提取方法。

#### 1. PCA

PCA 是一种无监督线性子空间特征提取方法。从线性代数的角度来看, PCA 的目的是通过线性变换 (或映射) 寻找一组最优的单位正交基, 用这组正交基的线性组合来重建原始数据, 使得重建后的数据和原始数据的均方误差最小。PCA 计算步骤如下。

(1) 计算协方差矩阵  $\mathbf{C}$ , 即

$$\mathbf{C} = 1/N \sum_{i=1}^N (\mathbf{x}_i - \mathbf{u})(\mathbf{x}_i - \mathbf{u})^T \quad (6-1)$$

其中,  $\mathbf{u}$  为所有样本的均值, 即  $\mathbf{u} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$ 。

(2) 对  $\mathbf{C}$  进行对角化, 得

$$\mathbf{C} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \quad (6-2)$$

其中,  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ ,  $\lambda_i$  为  $\mathbf{C}$  的特征值, 按照从大到小的顺序排列;

$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$ ,  $\mathbf{v}_i$  为特征值  $\lambda_i$  对应的特征向量, 满足  $\mathbf{V} \mathbf{V}^T = \mathbf{I}$ ,  $\mathbf{I}$  为单位矩阵。在这里, 特征向量  $\mathbf{v}_i$  称为主元 (Principle Component, PC),  $\mathbf{v}_1$  为最大特征值对应的特征向量, 因此又称为第一主元 (PC1), 其对应为投影方差最大的方向。图 6.3 中的实线为椭圆的长轴, 对应为投影方差最大的方向, 为第一主元。

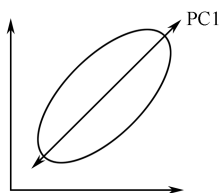


图 6.3 PCA 示意图

(3) 取最大的  $m'$  个特征值对应的特征向量作为子空间的基底,  $\mathbf{V}_{m'} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}]$ , 那么, 数据  $\mathbf{x}_i$  在  $m'$  个基底上投影为  $\mathbf{y}_i = \mathbf{V}_{m'}^T \mathbf{x}_i$ , 利用  $\mathbf{y}_i$  重建  $\mathbf{x}_i$ , 可以用  $\hat{\mathbf{x}}_i = \mathbf{V}_{m'} \mathbf{y}_i$  得到。

PCA 中主元的重要性通常是由其对应的特征值大小来确定的, 特征值越大, 其重要性越大。对于具体应用, 确定最佳主元的个数目前有两种常用的标准: ① 设一阈值  $\sigma_1$ ,  $\lambda_{\max}$  为最大的特征值, 保留所有满足  $\lambda_i / \lambda_{\max} > \sigma_1$  条件的特征值  $\lambda_i$  对应的主元; ② 设一阈值  $\sigma_2$ , 满足条件  $\sum_{i=1}^{m'} \lambda_i / \sum_{i=1}^m \lambda_i > \sigma_2$  的最小  $m'$  值即为最佳主元的个数。



现有的许多子空间特征提取方法都是 PCA 方法的变形和改进, 如对称 PCA (Symmetrical PCA, SPCA)<sup>[7]</sup>、单元 PCA (Modular PCA, mPCA)<sup>[8]</sup>、子模式 PCA (Sub-pattern PCA, SpPCA)<sup>[9]</sup>、自适应加权子模式 PCA (Adaptively Weighted Sub-pattern PCA, AwSpPCA)<sup>[10]</sup>等。SPCA 方法利用人脸的镜像对称性, 在镜像奇偶对称人脸图像上执行 PCA 方法。标准的 PCA 方法利用图像的全局信息, 在外界条件变化的情况下, PCA 的性能大大降低。mPCA 将图像分成大小相同的子图像, 然后在所有子图像形成的图像集上执行 PCA 方法, 在 mPCA 中, 外界条件的变化仅仅影响某些子图像而不会影响整个图像, 而 mPCA 忽略了子图像之间的空间关系。SpPCA 将图像分成大小相同的子图像, 相同位置的子图像形成的图像集称为子模式, 分别在每一个子模式上执行 PCA 方法, 然后将子模式得到的投影向量组成最终的投影向量, 但其并未考虑不同子模式对分类的贡献。在 AwSpPCA 方法中, 自适应确定不同子模式对分类的贡献。

## 2. LDA

与 PCA 方法不同, LDA 是一种有监督的线性子空间特征提取方法。LDA 具有较强的鉴别能力, 其目标是寻找一个变换矩阵  $\mathbf{V}$  将高维数据映射到低维空间内, 使得类内离差度最小, 类间离差度最大, 即在低维空间内同类别数据尽量靠近而不同类别数据尽量分离。LDA 分别采用类内离差矩阵  $\mathbf{S}_w$  和类间离差矩阵  $\mathbf{S}_b$  来刻画数据的聚合和分离程度, 其定义为

$$\mathbf{S}_w = \sum_{i=1}^M \sum_{j=1}^{n_i} (\mathbf{x}_i^j - \mathbf{u}_i)(\mathbf{x}_i^j - \mathbf{u}_i)^T \quad (6-3)$$

$$\mathbf{S}_b = \sum_{i=1}^M n_i (\mathbf{u}_i - \mathbf{u})(\mathbf{u}_i - \mathbf{u})^T \quad (6-4)$$

其中,  $\mathbf{u}$  为所有样本的均值;  $\mathbf{u}_i$  为第  $i$  类样本均值;  $n_i$  为第  $i$  类样本的个数。

LDA 采用 Fisher (费舍尔) 准则函数作为目标函数, Fisher 准则函数定义为

$$J(\mathbf{V}) = \arg \max_{\mathbf{V}} \frac{\text{trace}(\mathbf{V}^T \mathbf{S}_b \mathbf{V})}{\text{trace}(\mathbf{V}^T \mathbf{S}_w \mathbf{V})} \quad (6-5)$$

如果  $\mathbf{S}_w$  为非奇异矩阵, 式 (6-5) 的求解转换为式 (6-6) 对应的广义特征值和特征向量的求解问题, 即

$$\mathbf{S}_b \mathbf{v} = \lambda \mathbf{S}_w \mathbf{v} \quad (6-6)$$

假设将求出的特征值  $\lambda$  按照从大到小的顺序排列, 即有  $\lambda_i \geq \lambda_{i+1}$ , 则最优变换矩阵 (也称最优投影矩阵)  $\mathbf{V}$  由最大的  $m'$  个非零特征值对应的特征向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}$  组成。

在现实的应用中, 样本的个数远小于样本的维数,  $\mathbf{S}_w$  通常为奇异矩阵, 不能直接采用式 (6-6) 进行求解。为此, Swets 等<sup>[11]</sup>提出 PCA 与 LDA 相结合的特征提取方法, 即先利用 PCA 方法得到原始数据的低维子空间, 同时保证在该子空间内的  $\mathbf{S}_w$  为

非奇异矩阵, 然后在此基础上执行 LDA 方法。此外, 还有许多 LDA 的改进方法, 如直接 LDA<sup>[12,13]</sup>、直接加权 LDA<sup>[14]</sup>、零空间 LDA<sup>[15]</sup>、双空间 LDA<sup>[16]</sup>、规则化鉴别分析<sup>[17]</sup>、通用奇异值分解<sup>[18,19]</sup>等。

### 3. 其他方法

其他线性方法还包括典型相关分析 (Canonical Correlation Analysis, CCA)<sup>[20,21]</sup>、独立成分分析 (Independent Component Analysis, ICA)<sup>[22]</sup>、非负矩阵分解 (Non-negative Matrix Factorization, NMF)<sup>[23]</sup>等。

## 6.2.2 核方法

传统的 PCA、LDA 等方法本质上是线性的, 对于高度复杂非线性分布的数据的分类问题不能取得令人满意的结果, 通常采用核方法对线性方法进行扩展来处理非线性分布的数据。

### 1. 核方法的基本理论

核方法最初应用于将线性支持向量机 (SVM) 推广至非线性支持向量机<sup>[24,25]</sup>。核方法的基本思想就是通过非线性映射  $\phi: \mathbf{x} \in \mathbf{R}^m \rightarrow \phi(\mathbf{x}) \in H$ , 把在原始空间内线性不可分的样本映射到高维希尔伯特空间 (或称为核空间)  $H$ , 在空间  $H$  内样本线性可分, 如图 6.4 所示。

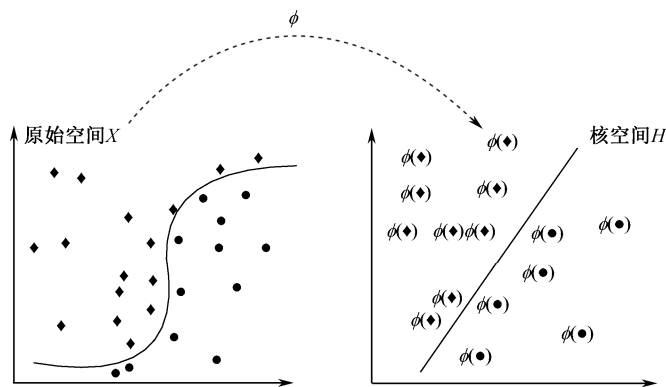


图 6.4 核方法示意图

为了方便描述与分析, 先给出一些重要的定义和定理。

**定义 6-1[核函数]** 核函数为定义在  $\mathbf{R}^m \times \mathbf{R}^m$  上的一个函数  $k: \mathbf{R}^m \times \mathbf{R}^m \rightarrow \mathbf{R}^1$ , 该函数对于任意  $\mathbf{x}, \mathbf{z} \in \mathbf{R}^m$ , 满足

$$k(\mathbf{x}, \mathbf{z}) = \langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle \quad (6-7)$$

其中,  $\phi$  为从  $\mathbf{R}^m$  到核空间  $H$  的非线性映射, 即  $\phi: \mathbf{x} \in \mathbf{R}^m \rightarrow \phi(\mathbf{x}) \in H$ ;  $\langle \cdot \rangle$  为内积运算。

**定义 6-2[核矩阵]** 对于给定的函数  $k: \mathbf{R}^m \times \mathbf{R}^m \rightarrow \mathbf{R}^1$  和  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N \in \mathbf{R}^m$ , 如果矩阵  $\mathbf{K} \in \mathbf{R}^{N \times N}$  的元素满足  $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ , 则称  $\mathbf{K}$  为关于  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  的核矩阵, 也称为 Gram 矩阵。

**定义 6-3[半正定矩阵]** 对于给定的矩阵  $\mathbf{K}$ , 如果对于任意的非零向量  $\mathbf{v} \in \mathbf{R}^N$ , 均有  $\mathbf{v}^T \mathbf{K} \mathbf{v} \geq 0$ , 则矩阵  $\mathbf{K}$  称为半正定矩阵。

**定理 6-1[Mercer 定理]** 对于给定的函数  $k: \mathbf{R}^m \times \mathbf{R}^m \rightarrow \mathbf{R}^1$ ,  $k$  为一个有效核 (也称为 Mercer 核函数) 的充分必要条件为关于  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  的 Gram 矩阵为对称的半正定矩阵。

常用的 Mercer 核函数有多项式核函数、高斯核函数 (径向基函数)、Sigmoid 核函数。它们分别定义如下。

(1) 多项式核函数的定义为

$$k(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x}, \mathbf{z} \rangle + 1)^p \quad (6-8)$$

其中,  $p$  为多项式阶数, 特别地, 当  $p=1$  时, 也称为线性核函数。

(2) 高斯核函数的定义为

$$k(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}\right) \quad (6-9)$$

(3) Sigmoid 核函数的定义为

$$k(\mathbf{x}, \mathbf{z}) = \tanh(v \langle \mathbf{x}, \mathbf{z} \rangle + c) \quad (6-10)$$

通过以上的分析可知, 我们不需要知道非线性映射  $\phi$  的具体形式, 并且可以利用 Mercer 核函数计算核空间  $H$  内两个向量的内积, 这样尽管核空间  $H$  的维数增加很多甚至达到无穷维, 但问题的计算复杂度并没有因此而增加多少, 且与核空间的维数无关。

对于一个直接处理输入数据空间的算法, 可以通过核技巧扩展为相应的核方法, 从而提高处理非线性输入数据的能力。图 6.5 显示了核方法处理的过程, 主要包括两个模块: 一是通过使用核函数将输入空间的数据隐式地映射到非线性的高维特征空间中; 二是在高维特征空间内使用线性分析方法。这两个模块相对独立, 通过核矩阵将两个模块连接起来<sup>[26]</sup>。

从图 6.5 中可知, 针对不同的应用背景, 可以对核方法和线性子空间特征提取方法进行单独设计, 从而得到多种不同的基于核方法的非线性特征提取方法。

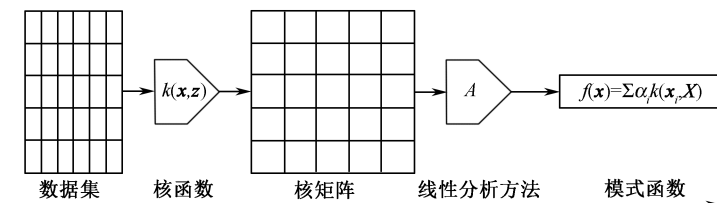


图 6.5 核方法处理过程示意图

## 2. 典型的核方法

传统的线性子空间特征提取方法可以利用核技巧得到其相应的非线性特征提取方法，如核主元成分分析（Kernel PCA，KPCA）<sup>[27]</sup>、核线性鉴别分析（Kernel LDA，KDA）<sup>[28]</sup>、核典型相关分析（Kernel CCA，KCCA）<sup>[29]</sup>、核独立成分分析（Kernel ICA，KICA）<sup>[30]</sup>等。

### 6.2.3 流形方法

流形（manifold）是局部具有欧氏空间性质的空间。流形学习就是从高维数据中恢复低维流形结构，即找到高维空间中的低维流形，并求出相应的嵌入映射<sup>[31]</sup>。图与流形有很多相近的性质，它们都可以嵌入到欧氏空间中，所以在很多情况下可以利用图来逼近流形，并采用图的理论求解低维嵌入。图中边的权值反映了图的拓扑性质，因此如果数据集足够大，噪声足够小，合理确定图中边的权值可以充分逼近嵌入流形<sup>[32]</sup>。2000 年，《科学》上发表了两篇基于流形学习的非线性特征提取方法的文章，分别为 Rowe 等人提出的局部线性嵌入（LLE）算法和 Tenenbaum 等人提出的拉普拉斯映射（LE）算法，自此基于流形学习的子空间特征提取方法的研究得到了迅猛发展。下面简单介绍 4 种经典的基于流形的特征提取方法。

#### 1. 等距映射（ISOMAP）<sup>[33]</sup>

ISOMAP 是建立在多维尺度变换（MDS）<sup>[34]</sup>基础上的一种全局保持流形学习方法，二者区别在于：MDS 全局保持两点之间的欧式距离，而 ISOMAP 力求保持两点之间的测地距离。ISOMAP 的关键是计算两点间的测地距离，根据数据点之间的近邻关系，构造邻接图，对于互为近邻的两点间的测地距离为两点的欧式距离，否则用邻接图中两点的最短路径距离来逼近。

ISOMAP 算法主要分为 4 个步骤。

（1）对于  $N$  个数据点的数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ，计算任意两个数据点之间的欧式距离， $d(i, j)$  表示  $\mathbf{x}_i$  和  $\mathbf{x}_j$  的欧式距离。

(2) 建立近邻图  $\mathbf{G}$ 。 $\mathbf{G}$  中的顶点与数据集中的数据点一一对应, 如果  $\mathbf{x}_i$  和  $\mathbf{x}_j$  的欧式距离小于某一个阈值  $\varepsilon$  ( $\varepsilon$  方法), 或者  $\mathbf{x}_i$  为与  $\mathbf{x}_j$  距离最近的  $k$  个数据点之一, 反之亦然 ( $k$  近邻法), 则  $\mathbf{x}_i$  和  $\mathbf{x}_j$  互为近邻点, 并且用一条边相连。图  $\mathbf{G}$  边的权值  $d_{\mathbf{G}}(i, j)$  定义为

$$d_{\mathbf{G}}(i, j) = \begin{cases} d(i, j), & \mathbf{x}_i \text{ 与 } \mathbf{x}_j \text{ 互为近邻点} \\ \infty, & \text{其他} \end{cases} \quad (6-11)$$

(3) 利用 Dijkstra 算法计算任意两个点之间的最短路径距离, 对于  $p=1, 2, \dots, N$ , 进行下面的替换运算。

$$d_{\mathbf{G}}(i, j) = \min \{d_{\mathbf{G}}(i, j), d_{\mathbf{G}}(i, p) + d_{\mathbf{G}}(p, j)\} \quad (6-12)$$

最终得到的矩阵  $\mathbf{D}_{\mathbf{G}} = [d_{\mathbf{G}}(i, j)]$  为最短路径距离矩阵。

(4) 计算数据集  $\mathbf{X}$  的低维映射  $\mathbf{Y}$ , 最小化式 (6-13) 对应的目标函数 (也称目标公式)。

$$E(\mathbf{Y}) = \|\tau(\mathbf{D}_{\mathbf{G}}) - \tau(\mathbf{D}_{\mathbf{Y}})\|_{L^2} \quad (6-13)$$

其中,  $\mathbf{D}_{\mathbf{Y}} = [d_{\mathbf{Y}}(i, j)] = [\|\mathbf{y}_i - \mathbf{y}_j\|_{L^2}]$  为欧式距离矩阵;  $\|\mathbf{A}\|_{L^2} = \sqrt{\sum_{i,j} A_{ij}^2}$  表示矩阵  $\mathbf{A}$  的 2 阶范数;  $\tau$  则是将距离转换为内积运算的操作, 令  $\mathbf{H} = \mathbf{I} - 1/N\mathbf{e}\mathbf{e}^T$ ,  $\mathbf{S}_{ij} = \mathbf{D}_{\mathbf{G}}^2(i, j)$ , 则  $\tau(\mathbf{D}_{\mathbf{G}}) = -\mathbf{H}\mathbf{S}\mathbf{H}/2$ 。

图 6.6 为 Swiss Roll 数据集上的实验结果。对于图 6.6 (a) 中 Swiss Roll 空间中的任意两个点 (圆圈表示), 虚线表示两点之间的欧式距离, 实线表示两点之间的测地距离; 图 6.6 (b) 中的实线表示两点之间的最短路径连接; 图 6.6 (c) 为 ISOMAP 算法得到的二维嵌入, 虚线表示二维空间内两点之间的测地距离, 而实线表示二维空间内两点之间的最短路径连接。从图 6.6 中可以看出, 利用近邻图能够较好地逼近数据的流形结构, 同时 ISOMAP 算法在降维的过程中能够很好地保持数据的流形结构。

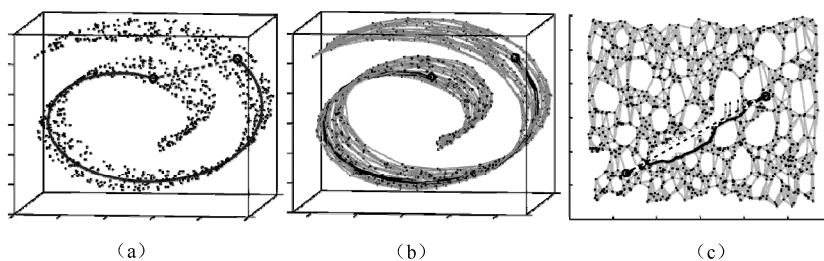


图 6.6 Swiss Roll 数据集上的实验结果

## 2. 局部线性嵌入 (LLE) [35]

LLE 算法是一种通过局部线性关系的组合来揭示全局非线性结构的非线性特征提取方法。LLE 假设采样数据所在的低维流形在局部是线性的, 即每个数据点可以通过

它的近邻点线性表示, 在实现中数据点用其近邻点的加权平均来表示, 其目标是把数据从高维空间映射到一个低维空间, 使得低维空间内数据点的近邻点的加权平均表示与在原始高维空间一致。

LLE 算法主要分为 3 步, 如图 6.7 所示。

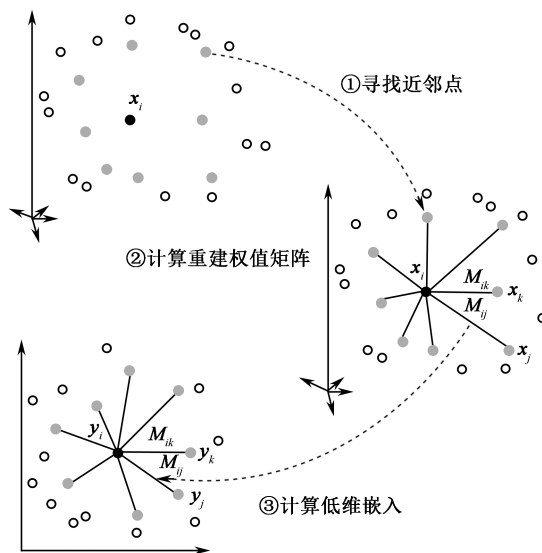


图 6.7 LLE 算法的流程

(1) 寻找近邻点。对于  $N$  个数据点的数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i \in \mathbf{R}^m$ , 利用  $k$  近邻方法计算每个数据点的近邻点集合, 记作  $N_k(\mathbf{x}_i)$ 。

(2) 计算重建权值矩阵  $\mathbf{M}$ 。对于每个数据点  $\mathbf{x}_i$  和它的近邻点集合  $N_k(\mathbf{x}_i)$ , 计算  $\mathbf{x}_i$  与近邻点之间的重建权值  $\mathbf{M}_{ij}$ , 权值  $\mathbf{M}_{ij}$  通过极小化重建误差来实现, 即

$$J(\mathbf{M}) = \min \sum_i \left\| \mathbf{x}_i - \sum_{\mathbf{x}_j \in N_k(\mathbf{x}_i)} \mathbf{M}_{ij} \mathbf{x}_j \right\|^2 \quad (6-14)$$

其中, 权值  $\mathbf{M}_{ij}$  表示近邻点  $\mathbf{x}_j$  对于数据点  $\mathbf{x}_i$  重建的贡献, 有  $\sum_{\mathbf{x}_j \in N_k(\mathbf{x}_i)} \mathbf{M}_{ij} = 1$ 。

(3) 计算低维嵌入。设  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$ ,  $\mathbf{y}_i \in \mathbf{R}^{m'}$  为数据集  $\mathbf{X}$  的低维嵌入, LLE 要求在低维空间内能够保持高维空间内数据之间的重建关系, 其目标公式使得低维空间内的重建误差极小化, 即

$$J(\mathbf{Y}) = \min \sum_i \left\| \mathbf{y}_i - \sum_j \mathbf{M}_{ij} \mathbf{y}_j \right\|^2 \quad (6-15)$$

对于 Swill Roll 数据集, LLE 算法的特征提取结果如图 6.8 所示。在图 6.8 中, (a) 为 Swill Roll 数据集, (b) 为从 (a) 中提取的样本点 (三维), (c) 为利用 LLE 算法

对 (b) 中的样本点进行特征提取得到的二维空间。从图 6.8 中可以看出, 在得到的二维空间内很好地保持了原有数据的近邻关系。

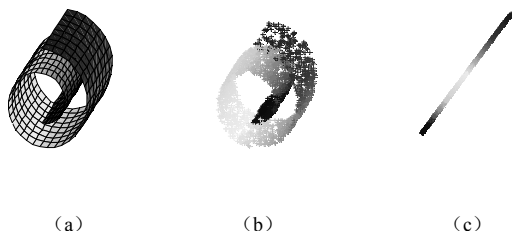


图 6.8 LLE 算法的降维结果

### 3. 拉普拉斯映射 (LE) [36]

LE 可以视为 LLE 方法的一种变形, 其基本思想为在高维空间中距离很近的样本点投影到低维空间中仍保持近邻关系。其基本原理为: 如果在嵌入的低维流形上均匀采样, 那么流形上的拉普拉斯算子可以由图的拉普拉斯矩阵来逼近, 这样 LE 算法的求解可以转换为求解图的拉普拉斯矩阵的广义特征值问题。

对于数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i \in \mathbf{R}^m$ , LE 算法主要有 3 步。

(1) 构建近邻图  $\mathbf{G} = (\mathbf{V}, \mathbf{W})$ 。图  $\mathbf{G}$  中的顶点与数据点一一对应, 利用  $k$  近邻或  $\varepsilon$  近邻方法计算每一个数据点的近邻点, 如果两个数据点  $\mathbf{x}_i$  和  $\mathbf{x}_j$  互为近邻点, 则  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间有一条边相连。

(2) 计算近邻图的权值矩阵  $\mathbf{W}$ 。权值矩阵  $\mathbf{W}$  的设置通常有两种方法: ① 0-1 方法, 如果  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间有一条边相连, 则  $\mathbf{W}_{ij} = 1$ , 否则  $\mathbf{W}_{ij} = 0$ ; ② 高斯核方法, 如果  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间有一条边相连, 其边的权值  $\mathbf{W}_{ij}$  为利用高斯核度量的样本之间的相似度, 即  $\mathbf{W}_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ , 否则  $\mathbf{W}_{ij} = 0$ 。

(3) 计算低维嵌入。LE 算法在特征提取的过程中保持高维空间中数据点之间的近邻关系, 设数据集  $\mathbf{X}$  的低维嵌入  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ ,  $\mathbf{y}_i \in \mathbf{R}^{m'}$ , 其目标公式为

$$J(\mathbf{Y}) = \min \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \mathbf{W}_{ij} \quad (6-16)$$

令  $\mathbf{D}$  为一个对角矩阵, 其值为权值矩阵  $\mathbf{W}$  的每一行或每一列的数据元素之和, 即  $D_{ii} = \sum_j \mathbf{W}_{ij}$ ,  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  为近邻图的拉普拉斯矩阵, 则低维嵌入  $\mathbf{Y}$  为式 (6-17) 对应的最小  $m'$  个特征值对应的特征向量组成。

$$\mathbf{L}\mathbf{y} = \lambda \mathbf{D}\mathbf{y} \quad (6-17)$$

#### 4. 局部保持投影映射 (LPP) [37]

ISOMAP、LLE 和 LE 等基于流形的子空间特征提取方法能够很好地发现嵌入在高维数据空间中低维流形,但是对于新的测试样本不能得到其低维数据表示,为此何晓飞等人提出了基于流形的线性子空间特征提取算法 LPP,它是 LE 算法的线性逼近。

##### 1) LPP 算法的基本思想

在识别问题中,两个样本的距离越近,其相似度越大,那么同属一个类别的可能性就越大。LPP 算法的基本思想就是寻找一个投影矩阵  $V$  将高维空间  $\mathbf{R}^m$  中的样本集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  映射为低维空间  $\mathbf{R}^{m'}$  ( $m' < m$ ) 中的样本集  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ , 即  $\mathbf{y}_i = V^T \mathbf{x}_i, (i=1, 2, \dots, N)$ , 使得在  $\mathbf{R}^m$  空间内互为近邻的两点经  $V$  映射后在  $\mathbf{R}^{m'}$  空间中仍互为近邻。

LPP 算法的过程主要包括以下 3 个步骤。

(1) 类似于 LE 算法,构造近邻图。

(2) 类似于 LE 算法,确定权值矩阵。

(3) 低维特征映射:最优投影矩阵的求解可以转换为式 (6-18) 对应的广义特征向量的求解问题,即

$$XLX^T \mathbf{v} = \lambda XDX^T \mathbf{v} \quad (6-18)$$

假定  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}$  为式 (6-18) 最小的  $m'$  个特征值对应的特征向量,则最优投影矩阵  $V$  为

$$V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}] \quad (6-19)$$

样本集  $\mathbf{X}$  的低维特征表示  $\mathbf{Y}$  为

$$\mathbf{y}_i = V^T \mathbf{x}_i, \quad i=1, 2, \dots, N \quad (6-20)$$

##### 2) LPP 算法的特性

LPP 与 PCA 均为无监督线性子空间特征提取方法,但与 PCA 相比, LPP 具有以下 3 个方面的优势:①LPP 算法具有较强的鲁棒性;②LPP 算法具有鉴别能力;③LPP 算法能够保持数据的拓扑结构。下面通过 3 个例子说明。

##### (1) 鲁棒性

在图 6.9 中,“\*”代表样本数据点(二维),右上角的数据点为一噪声点,图 6.9 (a) 为 PCA 的处理结果,图 6.9 (b) 为 LPP 的处理结果,其中长线为最优的投影方向,短线为次优的投影方向。在图 6.9 (a) 中,左下角的点在长线对应的投影方向的投影变为了一个点,因此 PCA 算法对噪声点比较敏感,并没有找到正确的最佳投影方向。而 LPP 由于能够保持局部信息,对噪声点不太敏感,具有较强的鲁棒性。



### (2) 鉴别能力

在图 6.10 中, 两个圆代表不同的类别, 图 6.10 (a) 为 PCA 的处理结果, 图 6.10 (b) 为 LPP 的处理结果, 其中长线为最优的投影方向, 短线为次优的投影方向。从处理结果看, PCA 使得两个圆在长线投影方向上重合在一起, 而 LPP 能够很好地将两个类别圆分开, 具有一定的鉴别能力。

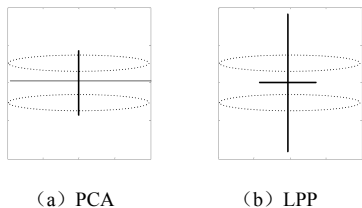


图 6.9 LPP 与 PCA 的鲁棒性比较

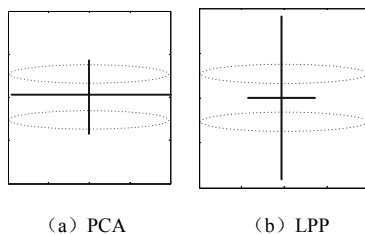


图 6.10 LPP 与 PCA 的鉴别能力比较

### (3) 拓扑结构保持能力

在图 6.11 中, 图 6.11 (a) 所示的样本点在三维空间内大概呈圆形分布, 图 6.11 (b) 为利用 PCA 降维后到二维空间, 样本点呈线状分布, 图 6.11 (c) 为利用 LPP 降维后得到的二维空间分布状况, 样本点仍然呈圆形分布, 可以看出 LPP 较 PCA 具有更好的拓扑结构保持能力。

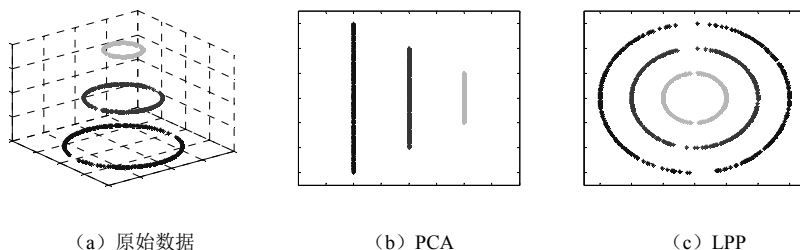


图 6.11 LPP 与 PCA 的拓扑结构保持能力比较

## 6.2.4 半监督方法

通常情况下有监督子空间特征提取方法的性能优于无监督子空间特征提取方法, 然而在现实应用中, 对样本进行类别标记费时费力, 无类别标记样本的获取相对比较容易, 因此, 半监督子空间特征提取方法成为一个研究热点, 如半监督子流形保持嵌入 (Semi-supervised Submanifold Preserving Embedding,  $S^3$ MPE)<sup>[38]</sup>、半监督鉴别分析 (Semi-supervised Discriminant Analysis, SDA)<sup>[39]</sup>、半监督局部费舍尔鉴别分析 (Semi-supervised Local Fisher Discriminant Analysis, SELF)<sup>[40]</sup> 和增强的 SELF

(Enhanced SELF, ESELF)<sup>[41]</sup>等,从大量无类别标记的样本和少量有类别标记的样本中进行特征提取得到相应的子空间。陈诗国等人<sup>[42]</sup>对一些半监督子空间特征提取方法的性能进行了比较,文献[43]、[44]从不同角度提出了通用的半监督子空间特征提取框架。

### 6.2.5 张量方法

前面提到的线性和非线性子空间特征提取方法都是基于向量的子空间特征提取方法,即在执行子空间方法之前,需要先将数据的特征信息转换为一维向量的表示形式,这样做的一个严重后果就是在转换的过程中丢失了数据行列的相关性。近年来,张量代数取得了重要的进展,基于二维矩阵或高维张量的子空间特征提取方法成为一个重要研究方向,如 2DPCA<sup>[45]</sup>、2DLDA<sup>[46]</sup>等。在这些方法中数据不需要转换为一维向量,而用二维矩阵的形式表示,但是在提取特征的过程中,仅考虑行与行之间的相关性而忽略了列与列之间的相关性,同时,得到的子空间的维数仍然非常高。为此,充分考虑了行与列两个方向上的相关性的双向特征提取方法[如 (2D)<sup>2</sup>PCA<sup>[47]</sup>、(2D)<sup>2</sup>LDA<sup>[48]</sup>和 TSA<sup>[49]</sup>等]相继提出。最近, Lu 等人提出了一个基于张量的多线性子空间学习的统一框架<sup>[50]</sup>,详细描述了多线性投影、多线性投影问题的求解、多线性学习方法的分类及应用,为设计基于张量的特征提取方法提供了重要的理论基础。

### 6.2.6 图嵌入框架

为了更好地揭示各种不同子空间特征提取方法之间的共同特征, Yan 等人提出图嵌入框架理论<sup>[51]</sup>,认为大多数子空间特征提取方法都可以统一到被称为“图嵌入”的框架中。图嵌入框架主要包括直接图嵌入及各种扩展形式,如线性化、核化和张量化。

#### 1. 直接图嵌入

在图嵌入框架下,大部分子空间特征提取方法都包含两类图:固有图 (intrinsic graph) 和惩罚图 (penalty graph)。

固有图  $G=(X, W)$  是一个无向带权图,图  $G$  中的顶点与样本一一对应,图中边的权值对应为样本点之间的相似度,用权值矩阵  $W \in \mathbf{R}^{N \times N}$  表示,通常有两种方法确定权值矩阵  $W$ : 高斯核函数方法和 0-1 方法 (具体见 6.2.3 小节中 LE 算法的介绍)。权值矩阵  $W$  描述了在子空间特征提取的过程中需要保持的某种统计或几何特性。固有图  $G$  的拉普拉斯矩阵  $L$  及对角矩阵  $D$  定义为

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad D_{ii} = \sum_j \mathbf{W}_{ij} \quad (6-21)$$

与固有图 $\mathbf{G}$ 一样, 惩罚图 $\mathbf{G}^p = (\mathbf{X}, \mathbf{W}^p)$ 也是一个带权的无向图, 图 $\mathbf{G}^p$ 中顶点与样本一一对应, 而权值矩阵 $\mathbf{W}^p$ 用来描述在子空间特征提取的过程中需要避免的某种统计或几何特性。惩罚图 $\mathbf{G}^p$ 的拉普拉斯矩阵 $\mathbf{L}^p$ 及对角矩阵 $\mathbf{D}^p$ 定义为

$$\mathbf{L}^p = \mathbf{D}^p - \mathbf{W}^p, \quad D_{ii}^p = \sum_j \mathbf{W}_{ij}^p \quad (6-22)$$

对于子空间特征提取问题, 需要构造一个固有图 $\mathbf{G}$ , 有选择性地构造惩罚图 $\mathbf{G}^p$ 。若 $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ 为样本集 $\mathbf{X}$ 的低维映射, 这里只讨论一维的情况, 即 $\mathbf{y}_i \in \mathbf{R}^1$ , 则直接图嵌入的目标函数定义为

$$J(\mathbf{Y}) = \min_{\mathbf{Y}^T \mathbf{B} \mathbf{Y} = \mathbf{I}} \sum_{i,j=1}^N \|\mathbf{y}_i - \mathbf{y}_j\|^2 \mathbf{W}_{ij} = \min_{\mathbf{Y}^T \mathbf{B} \mathbf{Y} = \mathbf{I}} \mathbf{Y}^T \mathbf{L} \mathbf{Y} \quad (6-23)$$

其中, 矩阵 $\mathbf{B}$ 为约束矩阵, 它可以是尺度归一化或另一个拉普拉斯矩阵, 即 $\mathbf{B} = \mathbf{L}^p = \mathbf{D}^p - \mathbf{W}^p$ , 这里 $\mathbf{W}^p$ 对应惩罚图 $\mathbf{G}^p$ 的权值矩阵。

利用直接图嵌入进行子空间特征提取可以很好地保持样本点高维空间内的相似度, 即对于两个差别较小的样本 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ , 二者的相似度较大(通常为正数), 为了使目标函数达到最小值,  $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 在子空间的投影 $\mathbf{y}_i$ 和 $\mathbf{y}_j$ 要离得比较近; 对于两个差别较大的样本 $\mathbf{x}_i$ 和 $\mathbf{x}_j$ , 二者的相似度较小(通常为负数), 为了使目标函数达到最小值,  $\mathbf{x}_i$ 和 $\mathbf{x}_j$ 在子空间的投影 $\mathbf{y}_i$ 和 $\mathbf{y}_j$ 要离得比较远。

## 2. 图嵌入的线性化、核化、张量化

直接图嵌入通过最优化式(6-23)对应的目标函数直接得到样本的低维投影, 但对于新的测试样本不能得到其低维投影, 这就是通常所讲的 Out-of-Sample 问题<sup>[52]</sup>, 因此需要将图嵌入框架进行线性化、核化或张量化来得到新的测试样本的低维投影。

### 1) 线性化 (linearization)

如果高维数据与其低维表示的映射函数 $F$ 采用线性函数, 即 $\mathbf{y}_i = \mathbf{v}^T \mathbf{x}_i$ , 其中 $\mathbf{v}$ 为变换向量(或投影向量), 则目标函数(6-23)变为

$$J(\mathbf{v}) = \min_{\substack{\mathbf{v}^T \mathbf{X} \mathbf{B} \mathbf{X}^T \mathbf{v} = \mathbf{I} \\ \text{or } \mathbf{v}^T \mathbf{v} = 1}} \sum_{i,j=1}^N \|\mathbf{v}^T \mathbf{x}_i - \mathbf{v}^T \mathbf{x}_j\|^2 \mathbf{W}_{ij} = \min_{\substack{\mathbf{v}^T \mathbf{X} \mathbf{B} \mathbf{X}^T \mathbf{v} = \mathbf{I} \\ \text{or } \mathbf{v}^T \mathbf{v} = 1}} \mathbf{v}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{v} \quad (6-24)$$

### 2) 核化 (kernelization)

利用核函数可以将线性算法推广到非线性情况, 即采用一个非线性映射将样本从原始空间映射到高维的希尔伯特空间 $H$ , 即 $\phi: \mathbf{x} \in \mathbf{R}^m \rightarrow \phi(\mathbf{x}) \in H$ , 然后在空间 $H$ 内执行线性操作。在空间 $H$ 内, 样本间的点积可以用样本间的核函数 $k$ 来代替, 有

$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ ，通过约束的变换方向  $\mathbf{v} = \sum_{i=1}^N a_i \phi(\mathbf{x}_i)$ ， $\boldsymbol{\alpha} = [a_1, a_2, \dots, a_N]^T$  为系数向量，目标函数 (6-23) 变为

$$J(\boldsymbol{\alpha}) = \min_{\substack{\boldsymbol{\alpha}^T \mathbf{K} \mathbf{L}^T \mathbf{K}^T \boldsymbol{\alpha} = 1 \\ \text{or } \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} = 1}} \sum_{i,j=1}^N \left\| \boldsymbol{\alpha}^T \mathbf{K}_i - \boldsymbol{\alpha}^T \mathbf{K}_j \right\|^2 W_{ij} = \min_{\substack{\boldsymbol{\alpha}^T \mathbf{K} \mathbf{L}^T \mathbf{K}^T \boldsymbol{\alpha} = 1 \\ \text{or } \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} = 1}} \boldsymbol{\alpha}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \boldsymbol{\alpha} \quad (6-25)$$

其中， $\mathbf{K}$  为核矩阵（也称为 Gram 矩阵），其元素表示为  $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ ，表示样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  在高维空间内的点积。

式 (6-23)、式 (6-24) 和式 (6-25) 的求解可以转换为式 (6-26) 对应的广义特征向量的求解，即

$$\tilde{\mathbf{L}} \mathbf{v} = \lambda \tilde{\mathbf{B}} \mathbf{v} \quad (6-26)$$

其中， $\tilde{\mathbf{L}}$  可以为  $\mathbf{L}$ 、 $\mathbf{X} \mathbf{L} \mathbf{X}^T$ 、 $\mathbf{K} \mathbf{L} \mathbf{K}$ ； $\tilde{\mathbf{B}}$  可以为  $\mathbf{L}$ 、 $\mathbf{B}$ 、 $\mathbf{K}$ 、 $\mathbf{X} \mathbf{B} \mathbf{X}^T$ 、 $\mathbf{K} \mathbf{B} \mathbf{K}$ 。

### 3) 张量化 (tensorization)

在上述图嵌入的线性化和核化中，样本  $\mathbf{x}_i$  用一个向量来表示，然而在现实生活中，样本的特征信息通常有一些特殊的结构，这种结构以二阶或更高阶的张量形式存在。例如，图像是一个二阶张量（矩阵），场景分析中视频序列可以看作一个三阶张量，如图 6.12 所示。为了应用线性化和核化方法，需要将张量数据展开为一个向量，这种转换大大增加了样本数据的维数，由于训练样本有限，通常会产生小样本问题，同时转换过程也破坏了数据的空间结构。

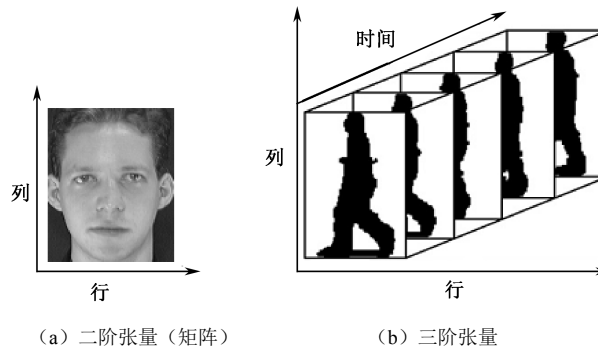


图 6.12 张量数据形式

在描述直接图张量化之前，简单介绍一下张量的相关概念和操作<sup>[53-55]</sup>。

$n$  阶张量  $\mathbf{X} \in \mathbb{R}^{m_1 \times m_2 \times \dots \times m_n}$  是一个高维数组，数据元素用  $X_{i_1, i_2, \dots, i_n}$  ( $1 \leq i_c \leq m_c, 1 \leq c \leq n$ ) 表示。

**定义 6-4[内积]** 两个  $n$  阶张量  $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m_1 \times m_2 \times \dots \times m_n}$  的内积为

$$\langle X, Y \rangle = \sum_{i_1=1, \dots, i_n=1}^{i_1=m_1, \dots, i_n=m_n} X_{i_1, \dots, i_n} Y_{i_1, \dots, i_n} \quad (6-27)$$

张量  $X$  的范数定义为  $\|X\| = \sqrt{\langle X, X \rangle}$ ，张量  $X$  与  $Y$  之间的距离表示为  $\|X - Y\|$ 。在二阶张量情况下，范数又称为 Frobenius 范数，记为  $\|X\|_F$ 。

**定义 6-5[k-模积]** 张量  $X$  与矩阵  $U \in \mathbf{R}^{m_k \times m'_k}$  的  $k$ -模积  $X \times_k U$  是一个  $m_1 \times m_2 \times \dots \times m_{k-1} \times m'_k \times m_{k+1} \times \dots \times m_n$  维张量，即

$$(X \times_k U)_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n} = \sum_{i_k=1}^{m_k} X_{i_1, \dots, i_{k-1}, i_k, i_{k+1}, \dots, i_n} \times U_{ij}, \quad j=1, 2, \dots, m'_k \quad (6-28)$$

**定义 6-6[k-模展开]** 通过固定张量第  $k$  维下标同时改变其他维下标的方法将一个张量  $X \in \mathbf{R}^{m_1 \times m_2 \times \dots \times m_n}$  展开为一个二维矩阵  $X^{(k)} \in \mathbf{R}^{m_k \times \prod_{i \neq k} m_i}$ ，记作

$$X \Rightarrow_k X^{(k)} \quad (6-29)$$

其中，

$$X^{(k)}_{i_k j} = X_{i_1, \dots, i_n}, \quad j = 1 + \sum_{p=1, p \neq k}^n (i_p - 1) \prod_{q=p+1, q \neq k}^n m_q \quad (6-30)$$

图 6.13 说明了一个三阶张量沿不同方向展开的结果。

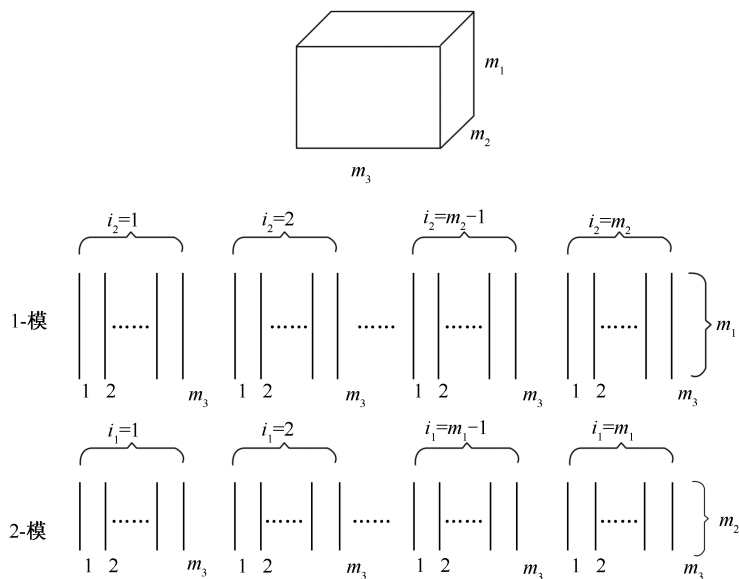


图 6.13 三阶张量  $k$ -模展开示意图

从图 6.13 可以看出，三阶张量 1-模展开的结果为一个  $m_1 \times (m_2 \times m_3)$  的二维数组，二维数组的行下标对应三阶张量的第一维下标，第二维下标对应三阶张量的第二维与第三维下标交替变化；三阶张量 2-模展开的结果为一个  $m_2 \times (m_1 \times m_3)$  的二维数组，二维数组的行下标对应三阶张量的第二维下标，第二维下标对应三阶张量的第一维与第三维下

标交替变化。

图嵌入的张量化可以形式化地描述如下：假设有一个含有  $N$  个  $n$  阶张量的样本集  $\{\mathbf{X}_i \in \mathbf{R}^{m_1 \times m_2 \times \cdots \times m_n}, i=1, 2, \cdots, N\}$ ，其目标就是寻找  $n$  个最优的投影矩阵  $(\mathbf{V}_i)_{i=1}^n \in \mathbf{R}^{m_i \times m'_i} (m'_i < m_i)$ ，将张量  $\mathbf{X} \in \mathbf{R}^{m_1 \times m_2 \times \cdots \times m_n}$  变为  $\mathbf{Y} \in \mathbf{R}^{m'_1 \times m'_2 \times \cdots \times m'_n}$ ，即

$$\mathbf{Y}_i = \mathbf{X}_i \times_1 \mathbf{V}_1 \times \cdots \times_n \mathbf{V}_n \quad (6-31)$$

在一维情况下，即投影矩阵  $(\mathbf{V}_i)_{i=1}^n \in \mathbf{R}^{m_i \times m'_i} (m'_i = 1)$ ，目标公式 (6-23) 变为

$$f(\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_n) = \min_{f(\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_n)=1} \sum_{i,j=1}^N \|\mathbf{X}_i \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times \cdots \times_n \mathbf{V}_n - \mathbf{X}_j \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times \cdots \times_n \mathbf{V}_n\|_{\mathbf{W}_{ij}} \quad (6-32)$$

如果约束矩阵  $\mathbf{B}$  为尺度归一化矩阵，那么

$$f(\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_n) = \sum_{i=1}^N \|\mathbf{X}_i \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times \cdots \times_n \mathbf{V}_n\|_{\mathbf{B}_{ii}} \quad (6-33)$$

如果约束矩阵  $\mathbf{B}$  为惩罚图对应的拉普拉斯矩阵，即  $\mathbf{B} = \mathbf{L}^p = \mathbf{D}^p - \mathbf{W}^p$ ，那么

$$f(\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_n) = \sum_{i,j=1}^N \|\mathbf{X}_i \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times \cdots \times_n \mathbf{V}_n - \mathbf{X}_j \times_1 \mathbf{V}_1 \times_2 \mathbf{V}_2 \times \cdots \times_n \mathbf{V}_n\|_{\mathbf{W}_{ij}^p}^p \quad (6-34)$$

在多数情况下，公式 (6-34) 没有通用的求解方法，通常采用迭代方法进行求解：对于每一个投影矩阵  $\mathbf{V}_o (o=1, 2, \cdots, n)$ ，假设  $(\mathbf{V}_1, \cdots, \mathbf{V}_{o-1}, \mathbf{V}_{o+1}, \cdots, \mathbf{V}_n)$  可知，令  $\mathbf{x}_i = \mathbf{X}_i \times_1 \mathbf{V}_1 \times \cdots \times_{o-1} \mathbf{V}_{o-1} \times_{o+1} \mathbf{V}_{o+1} \times \cdots \times_n \mathbf{V}_n$ ，目标公式 (6-34) 变为目标公式 (6-24)，这时可以采用线性化的方法求解出最优投影矩阵  $\mathbf{V}_o$ ，然后依次求解各个投影矩阵，直至收敛为止。张量化的特征提取方法在每一次迭代过程中需要考虑的特征维数远远小于线性化的特征提取方法，能够有效解决维数灾难问题，同时可以大大降低计算复杂度。

### 3. 图嵌入框架下的子空间特征提取方法

对于传统的子空间特征提取方法，大多都可以统一到图嵌入框架中，这些方法的区别主要在于固有图与惩罚图的构造及相应的权值矩阵  $\mathbf{W}$  和约束矩阵  $\mathbf{B}$  的定义。表 6-1 给出了前面提到的一些方法在图嵌入框架下权值矩阵  $\mathbf{W}$  和约束矩阵  $\mathbf{B}$  的定义。

表 6-1 图嵌入框架下的各种子空间特征提取方法的  $\mathbf{W}$  和  $\mathbf{B}$  的定义

方法	$\mathbf{W}$ 和 $\mathbf{B}$ 的定义
PCA/KPCA/2DPCA	$\mathbf{W}_{ij} = 1/N, \mathbf{B} = \mathbf{I}$
LDA/KDA/2DLDA	$\mathbf{W} = \sum_{c=1}^M \frac{1}{n_c} \mathbf{e}^c \mathbf{e}^{cT}, \mathbf{B} = \mathbf{I} - 1/N \mathbf{e} \mathbf{e}^T$
ISOMAP	$\mathbf{W}_{ij} = \tau(D_G)_{ij}, \mathbf{B} = \mathbf{I}$
LLE/NPE	$\mathbf{W} = \mathbf{M} + \mathbf{M}^T - \mathbf{M}^T \mathbf{M}, \mathbf{B} = \mathbf{I}$
LE/LPP	$\mathbf{W}_{ij} = \exp(-\ \mathbf{x}_i - \mathbf{x}_j\ ^2 / 2\sigma^2) [\mathbf{x}_i \in N_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_k(\mathbf{x}_i)], \mathbf{B} = \mathbf{D}$

在图嵌入框架的基础上，新的特征提取方法相继提出，如边界费舍尔分析(Maginal

Fisher Analysis, MFA)<sup>[56]</sup>、局部敏感鉴别分析 (Locality Sensitive Discriminant Analysis, LSDA)<sup>[57]</sup>、局部鉴别嵌入 (Local Discriminant Embedding, LDE)<sup>[58]</sup>等。

## 6.3 基于自适应近邻图嵌入的局部鉴别投影方法

### 6.3.1 方法提出的背景

Sugiyama 结合 LPP 与 LDA 的基本思想提出了局部费舍尔鉴别分析 (Local Fisher Discriminant Analysis, LFDA)<sup>[59]</sup>，然而 LFDA 在特征提取时仍有一些不足：①LFDA 未考虑不同类别数据间的近邻关系，相距较远的不同类别间的数据在类间离差度量时占据较大比重，以致在处理某些数据时得不到正确的最优投影方向<sup>[60]</sup>；②为描述数据的流形结构，LFDA 需要寻找样本的近邻点，近邻点个数的选择对最优投影方向的影响较大<sup>[61]</sup>。为了解决 LFDA 存在的不足，提出了基于自适应近邻图嵌入的局部鉴别投影 (neighborhood graph embedding based Local Adaptive Discriminant Projection, LADP) 方法<sup>[62]</sup>，同时考虑样本的类内和类间近邻关系，根据样本分布自适应确定近邻点的个数来消除其对最优投影子空间的影响，在得到的低维子空间内，使得相同类别的近邻点尽量靠近，不同类别的近邻点尽量分离。

### 6.3.2 LFDA

#### 1. 图嵌入框架下的 LDA 与 LFDA

图嵌入框架线性化可以将 LFDA 纳入其中，其权值矩阵  $\mathbf{W}$  和  $\mathbf{W}^p$  定义为

$$\begin{aligned} \mathbf{W}_{ij} &= \begin{cases} \mathbf{A}_{ij}(1/n_l), & l_i = l_j = l \\ 0, & l_i \neq l_j \end{cases} \\ \mathbf{W}_{ij}^p &= \begin{cases} \mathbf{A}_{ij}(1/N - 1/n_l), & l_i = l_j = l \\ 1/N, & l_i \neq l_j \end{cases} \end{aligned} \quad (6-35)$$

其中， $\mathbf{A}_{ij}$  是样本  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间的一种相似性度量， $\mathbf{x}_i$  和  $\mathbf{x}_j$  差别越小， $\mathbf{A}_{ij}$  值越大，反之越小，这里用高斯函数定义  $\mathbf{A}_{ij}$ ，即

$$\mathbf{A}_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2), & \mathbf{x}_i \in N_k(\mathbf{x}_j) \text{ 或 } \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \quad (6-36)$$

其中， $N_k(\mathbf{x}_j)$  表示与样本  $\mathbf{x}_j$  同类别的  $k$  个近邻点集合。从式 (6-35) 和式 (6-36) 可

以看出, LFDA 中的权值体现了鉴别信息与局部信息, 与  $\mathbf{x}_i$  和  $\mathbf{x}_j$  是否同类及是否相邻有关。

在式 (6-35) 权值矩阵的定义下,  $\mathbf{X}(\mathbf{D}-\mathbf{W})\mathbf{X}^T$  对应为 LFDA 算法中的局部类内离差矩阵, 由不同类别的所有样本点及同类别的近邻点决定,  $\mathbf{X}(\mathbf{D}^p-\mathbf{W}^p)\mathbf{X}^T$  对应为 LFDA 算法中的局部类间离差矩阵, 仅由同类别的近邻点决定。

## 2. LFDA 的不足

在 LFDA 中, 最优投影方向依赖于近邻点个数  $k$  的选择; 同时在计算类间离差度时, 未考虑不同类别数据之间的近邻关系, 对于某些数据, LFDA 得不到正确的最优投影方向。下面通过两个例子加以说明。

### 1) 未考虑不同类别数据之间的近邻关系对最优投影方向的影响

图 6.14 所示的数据产生自二元正态分布, 包含 “\*” 和 “+” 两类, 各有 100 个样本, 每一类又包含左、右两个聚类, 4 个聚类的均值分别为  $(-8, 4)$ 、 $(8, 4)$ 、 $(-8, -4)$ 、 $(8, -4)$ , 数据的协方差阵均为  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ , 图中的直线为 LFDA 方法在  $k=5$  时得到的最优投影方向。

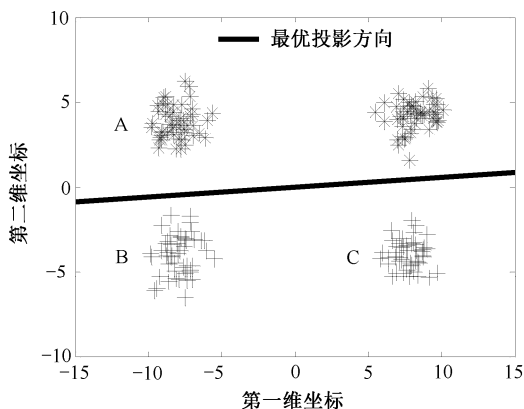


图 6.14 人工数据 1

在图 6.14 中, 聚类 A 与 B 的距离小于聚类 A 与 C 的距离, 显然最优投影方向应该为垂直方向, 但 LFDA 在类间离差度计算中所有不同类别的数据点间的距离的系数都是相同的, 均为  $1/N$ , 距离越大在类间离差度中所占的比重也越大。因为 A、C 之间的距离占主导地位, A、C 之间的距离在水平方向上的投影要大于在垂直方向上的投影, 因此得到的最优投影方向为水平方向。相反, 如果考虑不同类别数据之间的近邻关系, 这时 C 中的数据点就不会成为 A 中数据点的近邻点, 类间离差度仅由类间近



邻点决定,这时应该使得 A、B 两个聚类有最大分离程度,显然在垂直方向上满足要求。因此,在定义权值矩阵时需要考虑不同类别数据之间的近邻关系。

## 2) 近邻点个数 $k$ 的选择对最优投影方向的影响

图 6.15 所示的数据产生自二元正态分布,包含“\*”和“+”两类,每个类别包括 100 个样本点,类“\*”和类“+”的数据均值分别为  $(-3,0)$ 、 $(3,0)$ ,数据的协方差阵分别为  $\begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}$ 、 $\begin{bmatrix} 1 & 0 \\ 0 & 30 \end{bmatrix}$ 、 $\begin{bmatrix} 1 & 0 \\ 0 & 50 \end{bmatrix}$ 、 $\begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix}$ ,从图 6.14 (a) 到图 6.14 (d),水平方向上的方差不变,垂直方向上的方差依次增加,图中的直线为 LFDA 方法在  $k=5$  时得到的最优投影方向。

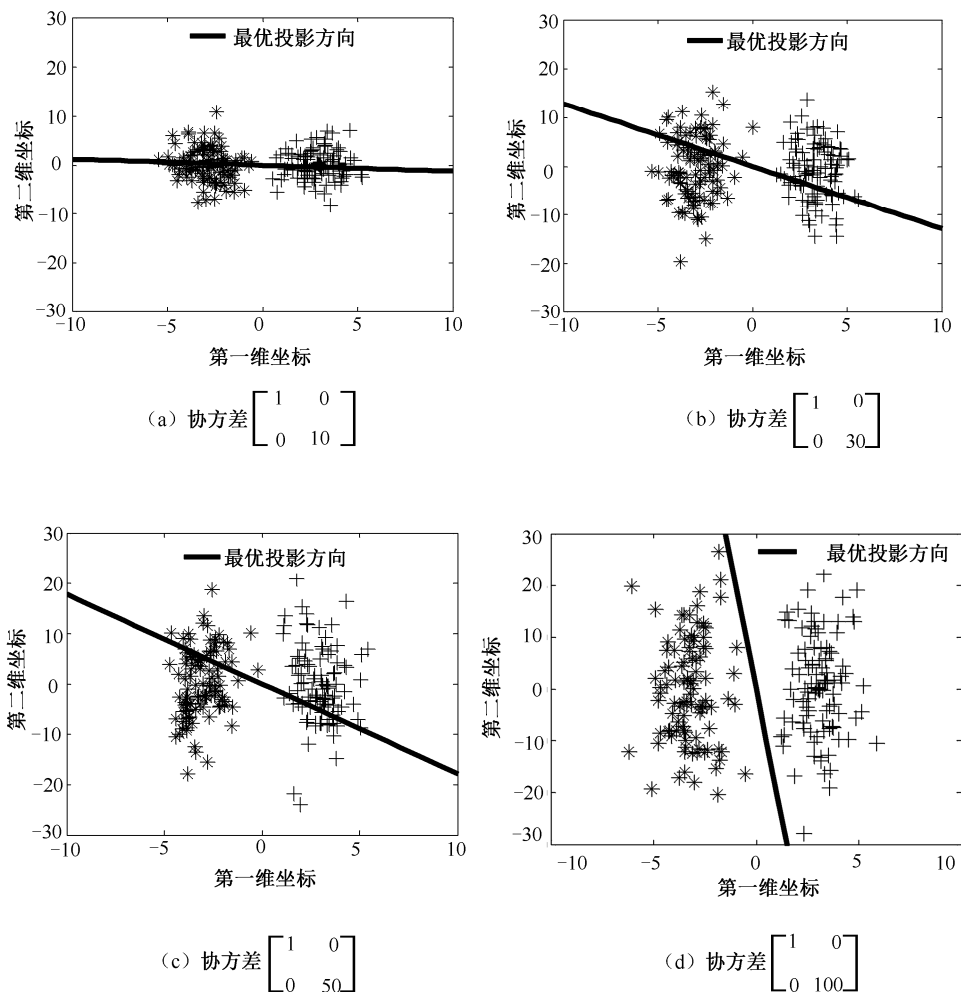


图 6.15 人工数据 2

很显然,对于图 6.15 中的人工数据,最优投影方向均为水平方向,但是数据点与其距离最远的数据点之间的距离随着垂直方向上方差的增加而不断增大。因为数据点与不同类别的数据点之间的权值均相等,因此与最远端的数据点之间的距离在类间离差度的计算中占有比较大的比重,又因为垂直方向上的方差远大于水平方向上的方差,因此投影方向随着方差的增加而逐渐向垂直方向靠近,这时就需要增加类内近邻点的个数(也就是  $k$  值)来抵消远端数据点在类间离差度中所占的比重。图 6.16 为对于图 6.15 (d) 中的数据,最优投影方向随近邻点个数  $k$  的变化情况。

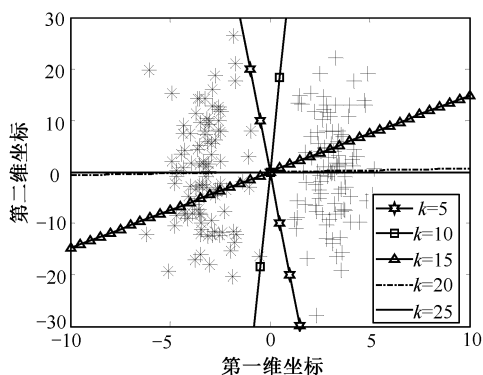


图 6.16 LFDA 在不同  $k$  值下的最优投影方向对比

从图 6.16 中可以看出,随着  $k$  值的增加,最优投影方向逐渐向水平方向靠近。因此,近邻点个数  $k$  对最优投影方向有较大的影响,需要根据样本的分布自适应确定样本之间的近邻关系。

### 6.3.3 LADP

基于 LFDA 的问题,这里提出了基于自适应近邻图嵌入的局部鉴别投影(LADP)方法。其步骤如下。

步骤 1: 根据样本分布特性及样本间相似度自适应计算样本类内及类间的近邻点,据此构造类内及类间近邻图。

步骤 2: 根据样本的类内及类间近邻点的个数与样本间相似度定义局部类内及类间离差矩阵中的权值矩阵,得到局部类内及类间离差矩阵。

步骤 3: 最大化局部类间离差矩阵与局部类内离差矩阵的迹比值,得到最优子空间。

#### 1. 基于样本分布的近邻图构造

根据图嵌入框架理论,为了更好地反映数据流形结构及鉴别信息,构造反映类内紧密性的类内近邻图  $\mathbf{G}$  (固有图)及反映类间分离性的类间近邻图  $\mathbf{G}^p$  (惩罚图)。在

近邻图  $G$  与  $G^p$  中, 顶点与样本一一对应, 顶点之间是否有边相连由顶点所对应的样本是否互为近邻点决定, 用  $l_i$  表示样本  $x_i$  的类别标号,  $N_w(x_i)$  表示与  $x_i$  具有相同类别标号的近邻点集合,  $N_b(x_i)$  表示与  $x_i$  具有不同类别标号的近邻点集合, 那么在近邻图  $G$  中,  $x_i$  与  $N_w(x_i)$  中的所有样本点有一条边相连, 同样, 在近邻图  $G^p$  中,  $x_i$  与  $N_b(x_i)$  中的所有样本点有一条边相连。为了克服近邻点个数  $k$  的影响, 根据样本分布自适应确定近邻点集合  $N_w(x_i)$  与  $N_b(x_i)$ 。

首先计算样本  $x_i$  与所有其他样本之间的平均相似度  $AS(x_i)$ 。

$$AS(x_i) = \frac{1}{N} \sum_{m=1}^N \exp\left(-\frac{\|x_i - x_m\|^2}{\beta}\right) \quad (6-37)$$

其中, 参数  $\beta$  取式 (6-38) 定义的所有样本之间的欧式距离的平均值, 即

$$\beta = \frac{1}{N^2} \sum_{i,j=1}^N \|x_i - x_j\|^2 \quad (6-38)$$

接着, 确定  $x_i$  同类别的近邻点集合  $N_w(x_i)$  及  $x_i$  不同类别的近邻点集合  $N_b(x_i)$ 。

$$N_w(x_i) = \left\{ x_j \mid l_j = l_i, \exp\left(-\frac{\|x_j - x_i\|^2}{\beta}\right) > AS(x_i) \right\} \quad (6-39)$$

$$N_b(x_i) = \left\{ x_j \mid l_j \neq l_i, \exp\left(-\frac{\|x_j - x_i\|^2}{\beta}\right) > AS(x_i) \right\} \quad (6-40)$$

根据式 (6-39) 和式 (6-40) 的定义,  $N_w(x_i)$  是所有与  $x_i$  同类别并且与  $x_i$  的相似度高于平均相似度的所有样本的集合,  $N_b(x_i)$  是所有与  $x_i$  不同类别并且与  $x_i$  的相似度高于平均相似度的所有样本的集合。图 6.17 为类内近邻点集合计算的示意图。

从图 6.17 中可以看出,  $x_1$  位于低密度区域, 平均相似度较低, 其近邻点个数也相对较少;  $x_2$  位于高密度区域, 平均相似度较高, 其近邻点个数也相对较多。每一个样本的近邻点的个数是不相同的, 取决于周围样本的分布。

## 2. 定义权值矩阵

设  $W$  和  $W^p$  分别为类内近邻图  $G$  及类间近邻图  $G^p$  的权值矩阵, 其值定义为

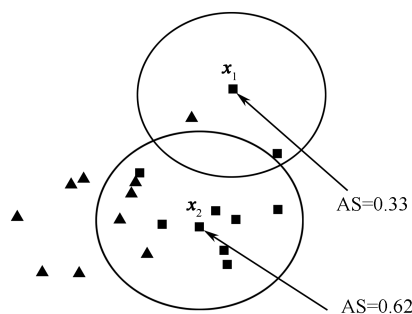


图 6.17 类内近邻点集合计算的示意图

$$\begin{aligned} \mathbf{W}_{ij} &= \begin{cases} \frac{A_{ij}}{k_w(i)}, & \mathbf{x}_i \in N_w(\mathbf{x}_j) \text{ 或 } \mathbf{x}_j \in N_w(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \\ \mathbf{W}_{ij}^p &= \begin{cases} A_{ij} \left[ \frac{1}{k_w(i) + k_b(i)} - \frac{1}{k_w(i)} \right], & \mathbf{x}_i \in N_w(\mathbf{x}_j) \text{ 或 } \mathbf{x}_j \in N_w(\mathbf{x}_i) \\ \frac{A_{ij}}{k_w(i) + k_b(i)}, & \mathbf{x}_i \in N_b(\mathbf{x}_j) \text{ 或 } \mathbf{x}_j \in N_b(\mathbf{x}_i) \\ 0, & \text{其他} \end{cases} \end{aligned} \quad (6-41)$$

其中,  $k_w(i)$  为样本点  $\mathbf{x}_i$  同类别的近邻点的个数, 其值为同类别的近邻点集合  $N_w(\mathbf{x}_i)$  中样本点的个数;  $k_b(i)$  为样本点  $\mathbf{x}_i$  不同类别的近邻点的个数, 其值为不同类别的近邻点集合  $N_b(\mathbf{x}_i)$  中样本点的个数。  $A_{ij}$  为式 (6-36) 定义的基于高斯核的样本间相似性度量值。

从式 (6-36) 可以看出, 权值矩阵  $\mathbf{W}$  和  $\mathbf{W}^p$  的值由样本的类内及类间近邻点的个数与样本间相似度决定。如果每个样本点的类内近邻点的个数等于类内样本点的个数, 类间近邻点的个数等于不同类别样本点的个数, 即有  $k_w(i) = n_i$ ,  $k_b(i) = N - n_i$ , 同时对于类间近邻点, 其相似性度量值  $A_{ij}$  取值为 1, 则式 (6-41) 演变为 LFDA 中式 (6-35) 定义的权值矩阵。

### 3. 最优低维嵌入

LADP 的目标就是在嵌入的低维子空间内, 既能保持数据的流形结构, 同时又能保持数据的鉴别信息, 也就是要使类内近邻图  $\mathbf{G}$  中关联的样本点在低维子空间内集中程度尽量大, 同时要使类间近邻图  $\mathbf{G}^p$  关联的样本点在低维子空间内的分离程度尽量大。这里, 我们利用局部类内离差矩阵  $\mathbf{S}_{IW}$  来反映  $\mathbf{G}$  关联的样本的聚合程度, 利用局部类间离差矩阵  $\mathbf{S}_{IB}$  来反映  $\mathbf{G}^p$  关联的样本的分离程度,  $\mathbf{S}_{IT}$  为局部总体离差矩阵。在式 (6-41) 定义的权值矩阵下, 有

$$\mathbf{S}_{IW} = \mathbf{X}(\mathbf{D} - \mathbf{W})\mathbf{X}^T = \mathbf{X}\mathbf{L}\mathbf{X}^T \quad (6-42)$$

$$\mathbf{S}_{IB} = \mathbf{X}(\mathbf{D}^p - \mathbf{W}^p)\mathbf{X}^T = \mathbf{X}\mathbf{L}^p\mathbf{X}^T \quad (6-43)$$

若  $\mathbf{y}_i$  为样本  $\mathbf{x}_i$  的低维映射, 则有  $\mathbf{y}_i = \mathbf{V}^T \mathbf{x}_i$ , 在嵌入的低维空间内的局部类内离差矩阵  $\mathbf{S}_{IW}^V$  和局部类间离差矩阵  $\mathbf{S}_{IB}^V$  与原始空间内的  $\mathbf{S}_{IW}$  和  $\mathbf{S}_{IB}$  的关系为

$$\mathbf{S}_{IW}^V = \mathbf{V}^T \mathbf{S}_{IW} \mathbf{V} \quad (6-44)$$

$$\mathbf{S}_{IB}^V = \mathbf{V}^T \mathbf{S}_{IB} \mathbf{V} \quad (6-45)$$

在嵌入的低维子空间内,  $\mathbf{S}_{IW}^V$  越小意味着同类样本的集中程度越大,  $\mathbf{S}_{IB}^V$  越大意味着不同类样本的分离程度越大, 因此 LADP 的目标公式为

$$\max_{V \in m \times m'} \frac{\text{trace}(\mathbf{S}_{\text{IB}}^V)}{\text{trace}(\mathbf{S}_{\text{IW}}^V)} = \max_{V \in m \times m'} \frac{\text{trace}(V^T \mathbf{S}_{\text{IB}} V)}{\text{trace}(V^T \mathbf{S}_{\text{IW}} V)} \quad (6-46)$$

式(6-46)的最大化问题可以转换为广义特征向量求解问题,即求

$$\mathbf{S}_{\text{IB}} \mathbf{v} = \lambda \mathbf{S}_{\text{IW}} \mathbf{v} \quad (6-47)$$

如果  $\mathbf{S}_{\text{IW}}$  为非奇异矩阵,令  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}$  为式(6-47)最大的  $m'$  个非零特征值对应的特征向量,则最优投影矩阵为  $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}]_{m \times m'}$ 。

由于样本的维数  $m$  远大于样本的个数  $N$ ,  $\mathbf{S}_{\text{IW}}$  通常为奇异矩阵,这时 LADP 算法陷入小样本问题。为此,需要使用 PCA 方法进行降维,使得在 PCA 子空间内  $\mathbf{S}_{\text{IW}}$  为非奇异矩阵,然后在 PCA 空间内执行 LADP 算法。

根据以上分析, LADP 算法的流程如下。

**输入:**  $M$  个类别的样本集  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i \in \mathbf{R}^m$ ,  $i = 1, 2, \dots, N$ 。

**执行过程:**

(1) 将样本投影到 PCA 主元空间内,令投影矩阵为  $V_{\text{PCA}}$ ,投影后的样本依然用  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  表示。

(2) 利用式(6-39)和式(6-40)根据样本分布及样本间相似度自适应确定样本的近邻点,在此基础上构造类内及类间近邻图,近邻图的权值由式(6-41)确定。

(3) 根据式(6-42)和式(6-43)计算局部类内离差矩阵  $\mathbf{S}_{\text{IW}}$  及局部类间离差矩阵  $\mathbf{S}_{\text{IB}}$ 。

(4) 求解式(6-47)所示的广义特征向量问题,得到最大的  $m'$  个非零特征值对应的特征向量组成的投影矩阵  $V^* = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m'}]$ ,最终的投影矩阵为  $V = V_{\text{PCA}} V^*$ 。

**输出:** 投影矩阵  $V$ 。

#### 4. 时间复杂度分析

LFDA 与 LADP 算法的时间复杂度主要由两个方面决定:①类内及类间近邻点的计算;②广义特征向量的求解。

在 LFDA 算法中,类内及类间近邻点计算的时间复杂度为  $O(mN^2 + kN^2)$ 。其中,  $O(mN^2)$  代表计算任意两个样本的欧式距离的时间复杂度;  $m$  为样本的维数;  $N$  为样本的个数;  $O(kN^2)$  代表寻找同类别  $k$  个近邻点的时间复杂度。又有局部类内及类间离差矩阵均为  $m \times m$  矩阵,求解广义特征向量的时间复杂度为  $O(m^3)$ 。因此, LFDA 算法的时间复杂度为  $O(mN^2 + kN^2 + m^3)$ 。由于  $k$  远小于样本维数  $m$  和样本个数  $N$ ,  $O(kN^2)$  的变化趋势远小于  $O(mN^2)$  与  $O(m^3)$ ,所以 LFDA 算法的时间复杂度近似于  $O(mN^2 + m^3)$ 。

LADP 与 LFDA 的区别在于自适应确定类内及类间近邻点,计算任意两个样本的相似度的时间复杂度为  $O(mN^2)$ ,得到每一个样本的平均相似度的时间复杂度为  $O(N^2)$ ,与平均相似度进行比较确定类内及类间近邻点集合的时间复杂度为  $O(N^2)$ ,求解广义特征向量的时间复杂度为  $O(m^3)$ ,所以 LADP 算法的时间复杂度为

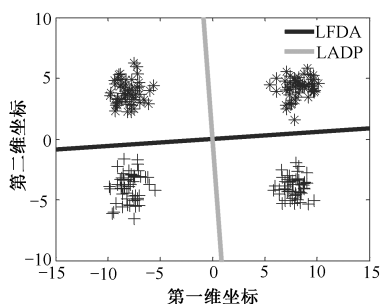


图 6.18 人工数据 1 下用 LFDA 与 LADP 得到的最优投影方向比较

$O(mN^2 + 2N^2 + m^3)$ 。由于  $O(2N^2)$  的变化趋势远小于  $O(mN^2)$  与  $O(m^3)$ ，所以 LADP 算法的时间复杂度近似于  $O(mN^2 + m^3)$ 。

因此，LADP 与 LFDA 的时间复杂度相当，仅由样本的个数  $N$  和样本的维数  $m$  决定。

## 5. LFDA 与 LADP 最优投影方向比较

图 6.18 和图 6.19 为 6.3.2 小节中提到的人工数据用 LFDA 与 LADP 得到的最优投影方向比较，图中的直线代表不同算法得到的最优投影方向。

在图 6.18 中，由于 LADP 仅考虑类间近邻点对类间离差度的影响，避免了远端聚类对最优投影方向的影响。在图 6.19 中，由于 LADP 根据样本分布自适应确定样本的近邻点，避免了近邻点个数选择对最优投影方向的影响。因此，对于 6.3.2 小节中提到的人工数据，相较于 LFDA，LADP 均能找到正确的最优投影方向。

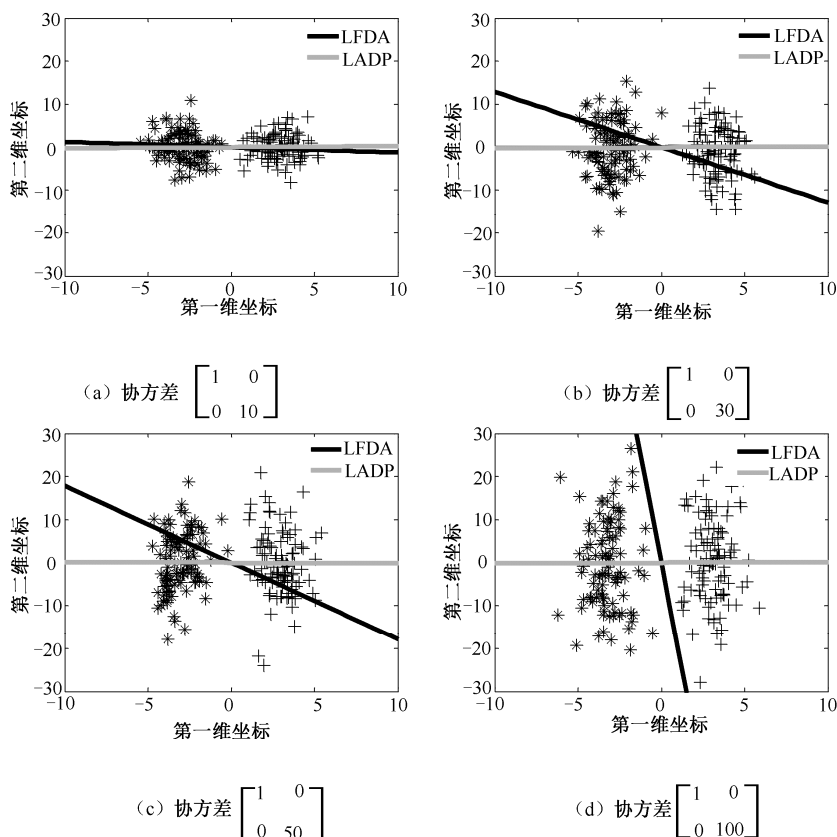


图 6.19 人工数据 2 下用 LFDA 与 LADP 得到的最优投影方向比较

## 6.4 基于对角图像的模糊线性鉴别分析

### 6.4.1 方法提出的背景

LDA 是一种有监督的子空间特征提取方法,其寻找一个最优投影矩阵,使得投影到低维空间的样本数据具有较好的可分离性,即使在投影空间中的不同类别样本间的差异尽可能大,同类别样本间的差异尽可能小。但 LDA 在对样本类别划分时采用“硬划分”,即认为某个样本要么属于某个类别,要么不属于某个类别。由于训练样本在获取的过程中受许多复杂条件的影响,例如,人脸图像往往受光照、表情、姿态等因素的影响,这样就不能简单地将样本归为某一类。文献[63]将模糊隶属的概念<sup>[64]</sup>与 LDA 相结合提出了模糊线性鉴别分析(Fuzzy Linear Discriminant Analysis, FLDA),取得了较好的效果。

同 LDA 一样,FLDA 也是基于向量的特征提取方法,即在进行特征提取之前,需先将图像数据的特征信息转换为一维向量,这种转换大大增加了样本数据的维数,由于训练样本有限,通常会产生小样本问题,同时转换过程也破坏了数据的空间结构。2DFLDA (Two Dimension FLDA)<sup>[65]</sup>直接应用于图像矩阵,提取图像行内的特征信息,但忽略了图像列内信息的变化。A2DFLDA (Alternative-2DFLDA)<sup>[66]</sup>同样也应用于图像矩阵,与 2DFLDA 不同,其提取图像列内的特征信息,忽略了行内信息的变化。张道庆等人提出将图像转换为对角图像,然后利用 PCA 方法对角图像进行特征提取,即对角 PCA (Diagonal PCA, DiaPCA)<sup>[67]</sup>,能够很好地提取出行内和列内的特征信息。随后,对角 LDA (Diagonal LDA, DiaLDA)<sup>[68]</sup>和对角 LPP (Diagonal LPP, DiaLPP)<sup>[69]</sup>等特征提取方法相继被提出。

在本节中,首先以引理的方式将 FLDA 方法统一到 6.2.6 小节提到的图嵌入框架中,即模糊隶属度矩阵与图嵌入框架中对应的固有图和惩罚图权值矩阵之间的对应关系;在此基础上,将对角图像与 FLDA 相结合提出了对角模糊线性鉴别分析(Diagonal FLDA, DiaFLDA)<sup>[70]</sup>。

### 6.4.2 FLDA

#### 1. 图嵌入框架下的 FLDA

假设有一个  $n$  个样本  $c$  个模糊类组成的样本集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ ,  $\mathbf{x}_i \in \mathbf{R}^D$ ,  $l_i \in L$

$= \{1, 2, \dots, c\}$  为样本  $\mathbf{x}_i$  的类别标号, 样本的平均值为  $\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ 。一个模糊  $c$  类划分用一个模糊隶属度矩阵  $\mathbf{U} = [u_{li}] (l=1, 2, \dots, c, i=1, 2, \dots, n)$  表示,  $u_{li}$  为样本  $\mathbf{x}_i$  对于类别  $l$  的隶属度, 满足  $\sum_{l=1}^c u_{li} = 1$ 。在此基础上, 类中心  $\mathbf{m}_l$  定义为

$$\mathbf{m}_l = \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j}{\sum_{k=1}^n u_{lk}}, \quad l=1, 2, \dots, c \quad (6-48)$$

模糊类内、类间及总体离差矩阵分别记作  $\mathbf{S}_{fw}$ 、 $\mathbf{S}_{fb}$  和  $\mathbf{S}_{ft}$ , 其定义为

$$\mathbf{S}_{fw} = \sum_{l=1}^c \sum_{i=1}^n u_{li} (\mathbf{x}_i - \mathbf{m}_l)(\mathbf{x}_i - \mathbf{m}_l)^T \quad (6-49)$$

$$\mathbf{S}_{fb} = \sum_{l=1}^c \sum_{i=1}^n u_{li} (\mathbf{m}_l - \mathbf{m})(\mathbf{m}_l - \mathbf{m})^T \quad (6-50)$$

$$\mathbf{S}_{ft} = \mathbf{S}_{fw} + \mathbf{S}_{fb} = \sum_{l=1}^c \sum_{i=1}^n u_{li} (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (6-51)$$

FLDA 就是要寻找一个最优的投影矩阵  $\mathbf{V}$ , 使得在低维空间内模糊类间离差矩阵  $\mathbf{S}_{fb}$  和模糊类内离差矩阵  $\mathbf{S}_{fw}$  的迹比值最大, 其目标公式为

$$\max_{\mathbf{V} \in D \times d} \frac{\text{trace}(\mathbf{V}^T \mathbf{S}_{fb} \mathbf{V})}{\text{trace}(\mathbf{V}^T \mathbf{S}_{fw} \mathbf{V})} \quad (6-52)$$

式 (6-52) 的最大化问题可以转换为广义特征向量求解问题, 即求

$$\mathbf{S}_{fb} \mathbf{v} = \lambda \mathbf{S}_{fw} \mathbf{v} \quad (6-53)$$

最优投影矩阵  $\mathbf{V}$  由式 (6-53) 最大的  $d$  个特征值对应的特征向量组成, 即  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$ 。

在 6.2.6 小节中提到大多数子空间特征提取方法都可以统一到图嵌入框架中, 各种特征提取方法的区别仅仅在于固有图与惩罚图的权值矩阵的选取。同样, FLDA 也能够统一到图嵌入框架中。在 Fisher 极大判别准则下, 6.2.6 小节中的图嵌入线性化框架等价的目标公式为

$$\max_{\mathbf{V} \in D \times d} \frac{\text{trace}[\mathbf{V}^T \mathbf{X}(\mathbf{D}^p - \mathbf{W}^p) \mathbf{X}^T \mathbf{V}]}{\text{trace}[\mathbf{V}^T \mathbf{X}(\mathbf{D} - \mathbf{W}) \mathbf{X}^T \mathbf{V}]} \quad (6-54)$$

其中,  $\mathbf{W}$  和  $\mathbf{W}^p$  为图嵌入框架中固有图和惩罚图对应的权值矩阵。下面通过引理 6-1 说明模糊隶属度矩阵  $\mathbf{U}$  与  $\mathbf{W}$  和  $\mathbf{W}^p$  之间的对应关系, 从而将 FLDA 统一到图嵌入框架中。

**引理 6-1** 在图嵌入框架下, 对于 FLDA, 式 (6-49) 和式 (6-50) 定义的  $\mathbf{S}_{fw}$  及  $\mathbf{S}_{fb}$  分别等价于式 (6-54) 中的  $\mathbf{X}(\mathbf{D} - \mathbf{W}) \mathbf{X}^T$  和  $\mathbf{X}(\mathbf{D}^p - \mathbf{W}^p) \mathbf{X}^T$ , 即  $\mathbf{S}_{fw}$  和  $\mathbf{S}_{fb}$  可以用点对表示形式, 即



$$\mathbf{S}_{\text{fw}} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\text{T}} = \mathbf{X}(\mathbf{D} - \mathbf{W})\mathbf{X}^{\text{T}} \quad (6-55)$$

$$\mathbf{S}_{\text{fb}} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij}^{\text{p}} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\text{T}} = \mathbf{X}(\mathbf{D}^{\text{p}} - \mathbf{W}^{\text{p}})\mathbf{X}^{\text{T}} \quad (6-56)$$

其中,  $\mathbf{W}_{ij} = \frac{\sum_{l=1}^c \frac{u_{li}u_{lj}}{\sum_{k=1}^n u_{lk}}}{\sum_{k=1}^n u_{lk}}$  和  $\mathbf{W}_{ij}^{\text{p}} = \frac{1}{n} - \mathbf{W}_{ij}$  分别表示固有图与惩罚图的权值矩阵。

证明:

$$\begin{aligned} \mathbf{S}_{\text{fw}} &= \sum_{l=1}^c \sum_{i=1}^n u_{li} \left( \mathbf{x}_i - \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j}{\sum_{k=1}^n u_{lk}} \right) \left( \mathbf{x}_i - \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j}{\sum_{k=1}^n u_{lk}} \right)^{\text{T}} \\ &= \sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^{\text{T}} - \sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j^{\text{T}}}{\sum_{k=1}^n u_{lk}} \end{aligned} \quad (6-57)$$

$\mathbf{S}_{\text{fw}}$  由两项组成, 对于  $\sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^{\text{T}}$  经过简单的变形, 变为

$$\sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^{\text{T}} = \sum_{i=1}^n \sum_{l=1}^c u_{li} \mathbf{x}_i \mathbf{x}_i^{\text{T}} = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^{\text{T}} \sum_{l=1}^c u_{li} \quad (6-58)$$

由于  $\sum_{l=1}^c u_{li} = 1$ , 所以有

$$\sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^{\text{T}} = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^{\text{T}} \quad (6-59)$$

对于  $\sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j^{\text{T}}}{\sum_{k=1}^n u_{lk}}$  经过简单的变形, 变为

$$\sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j^{\text{T}}}{\sum_{k=1}^n u_{lk}} = \sum_{i=1}^n \sum_{l=1}^c u_{li} \mathbf{x}_i \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j^{\text{T}}}{\sum_{k=1}^n u_{lk}} = \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^c \frac{u_{li}u_{lj}}{\sum_{k=1}^n u_{lk}} \mathbf{x}_i \mathbf{x}_j^{\text{T}} \quad (6-60)$$

由于

$$\sum_{j=1}^n \sum_{l=1}^c \frac{u_{li}u_{lj}}{\sum_{k=1}^n u_{lk}} = \sum_{l=1}^c \sum_{j=1}^n \frac{u_{li}u_{lj}}{\sum_{k=1}^n u_{lk}} = \sum_{l=1}^c u_{li} \sum_{j=1}^n \frac{u_{lj}}{\sum_{k=1}^n u_{lk}} = \sum_{l=1}^c u_{li} = 1 \quad (6-61)$$

所以, 由式 (6-59) 和式 (6-61), 有

$$\sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^T = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T = \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^c \frac{u_{li} u_{lj}}{\sum_{k=1}^n u_{lk}} \mathbf{x}_i \mathbf{x}_i^T \quad (6-62)$$

由式 (6-60) 和式 (6-62), 有

$$\begin{aligned} \mathbf{S}_{\text{fw}} &= \sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^T - \sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \frac{\sum_{j=1}^n u_{lj} \mathbf{x}_j^T}{\sum_{k=1}^n u_{lk}} \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^c \frac{u_{li} u_{lj}}{\sum_{k=1}^n u_{lk}} \mathbf{x}_i \mathbf{x}_i^T - \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^c \frac{u_{li} u_{lj}}{\sum_{k=1}^n u_{lk}} \mathbf{x}_i \mathbf{x}_j^T \end{aligned} \quad (6-63)$$

令  $\mathbf{W}_{ij} = \frac{\sum_{l=1}^c u_{li} u_{lj}}{\sum_{k=1}^n u_{lk}}$ ,  $\mathbf{S}_{\text{fw}}$  变为

$$\begin{aligned} \mathbf{S}_{\text{fw}} &= \sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{ij} \mathbf{x}_i \mathbf{x}_i^T - \sum_{i=1}^n \sum_{j=1}^n \mathbf{W}_{ij} \mathbf{x}_i \mathbf{x}_j^T \\ &= \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{x}_i \mathbf{x}_i^T + \mathbf{x}_j \mathbf{x}_j^T - \mathbf{x}_i \mathbf{x}_j^T - \mathbf{x}_j \mathbf{x}_i^T) \\ &= \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \end{aligned} \quad (6-64)$$

式 (6-55) 得证。

对于模糊总体离差矩阵  $\mathbf{S}_{\text{ft}}$ , 有

$$\begin{aligned} \mathbf{S}_{\text{ft}} &= \sum_{l=1}^c \sum_{i=1}^n u_{li} \left( \mathbf{x}_i - \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \right) \left( \mathbf{x}_i - \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j \right)^T \\ &= \sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \mathbf{x}_i^T - \sum_{l=1}^c \sum_{i=1}^n u_{li} \mathbf{x}_i \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^T \\ &= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T - \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n} \mathbf{x}_i \mathbf{x}_j^T \sum_{l=1}^c u_{li} \\ &= \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T - \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n} \mathbf{x}_i \mathbf{x}_j^T \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n} \mathbf{x}_i \mathbf{x}_i^T - \sum_{i=1}^n \sum_{j=1}^n \frac{1}{n} \mathbf{x}_i \mathbf{x}_j^T \\ &= \frac{1}{2} \sum_{i,j=1}^n \frac{1}{n} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \end{aligned} \quad (6-65)$$

由于  $\mathbf{S}_{\text{fb}} = \mathbf{S}_{\text{ft}} - \mathbf{S}_{\text{fw}}$ , 则有

$$\begin{aligned}
\mathbf{S}_{\text{fb}} &= \mathbf{S}_{\text{ft}} - \mathbf{S}_{\text{fw}} \\
&= \frac{1}{2} \sum_{i,j=1}^n \frac{1}{n} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\text{T}} - \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\text{T}} \\
&= \frac{1}{2} \sum_{i,j=1}^n \left( \frac{1}{n} - \mathbf{W}_{ij} \right) (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\text{T}}
\end{aligned} \tag{6-66}$$

令  $\mathbf{W}_{ij}^{\text{p}} = \frac{1}{n} - \mathbf{W}_{ij}$ ，式 (6-56) 得证。

**推论 6-1** FLDA 在图嵌入框架下的权值矩阵  $\mathbf{W}$  和  $\mathbf{W}^{\text{p}}$  与模糊隶属度矩阵  $\mathbf{U}$  具有如下的关系：

$$\mathbf{W} = \mathbf{A}^{\text{T}} \mathbf{A}, \quad \mathbf{W}^{\text{p}} = \frac{1}{n} \mathbf{I}_{n \times n} - \mathbf{A}^{\text{T}} \mathbf{A} \tag{6-67}$$

其中， $(\mathbf{A})_{l,i} = \frac{u_{li}}{\sqrt{\sum_{k=1}^n u_{lk}}} (l=1,2,\dots,c, i=1,2,\dots,n)$ ； $\mathbf{I}_{n \times n} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}_{n \times n}$ 。

## 2. 图嵌入框架下的 2DFLDA 与 A2DFLDA

与 FLDA 不同，2DFLDA 与 A2DFLDA 直接应用于图像矩阵进行特征提取，而不需要将图像按行或列展开为一个一维向量。

### 1) 2DFLDA

2DFLDA 可以提取出图像的行内信息，其形式化描述如下：假设有  $n$  个大小为  $\text{row} \times \text{col}$  的训练样本图像  $\mathbf{X}_i (i=1,2,\dots,n)$ ，寻找最优投影矩阵  $\mathbf{V} \in \mathbf{R}^{\text{col} \times d}$  使得  $\mathbf{X}_i$  的低维投影  $\mathbf{Y}_i = \mathbf{X}_i \mathbf{V}$ ， $\mathbf{Y}_i \in \mathbf{R}^{\text{row} \times d} (d < \text{col})$ ，使得在低维空间内模糊类内离差度最小的同时模糊类间离差度最大。在图嵌入框架下，2DFLDA 的模糊类内、类间离差矩阵  $\mathbf{S}_{\text{fw}}$  和  $\mathbf{S}_{\text{fb}}$  分别定义为

$$\mathbf{S}_{\text{fw}} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{X}_i - \mathbf{X}_j)^{\text{T}} (\mathbf{X}_i - \mathbf{X}_j) \tag{6-68}$$

$$\mathbf{S}_{\text{fb}} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij}^{\text{p}} (\mathbf{X}_i - \mathbf{X}_j)^{\text{T}} (\mathbf{X}_i - \mathbf{X}_j) \tag{6-69}$$

其中， $\mathbf{W}$  和  $\mathbf{W}^{\text{p}}$  为 FLDA 在图嵌入框架下的权值矩阵。

对于任意  $\mathbf{X}_i, \mathbf{X}_j \in \mathbf{R}^{\text{row} \times \text{col}}$ ，令  $\mathbf{X}_i^q$  表示  $\mathbf{X}_i$  的第  $q$  行，则有

$$\begin{aligned}
 (\mathbf{X}_i - \mathbf{X}_j)^T (\mathbf{X}_i - \mathbf{X}_j) &= \left( \begin{bmatrix} \mathbf{X}_i^1 \\ \mathbf{X}_i^2 \\ \vdots \\ \mathbf{X}_i^{\text{row}} \end{bmatrix} - \begin{bmatrix} \mathbf{X}_j^1 \\ \mathbf{X}_j^2 \\ \vdots \\ \mathbf{X}_j^{\text{row}} \end{bmatrix} \right)^T \left( \begin{bmatrix} \mathbf{X}_i^1 \\ \mathbf{X}_i^2 \\ \vdots \\ \mathbf{X}_i^{\text{row}} \end{bmatrix} - \begin{bmatrix} \mathbf{X}_j^1 \\ \mathbf{X}_j^2 \\ \vdots \\ \mathbf{X}_j^{\text{row}} \end{bmatrix} \right) \\
 &= \sum_{q=1}^{\text{row}} (\mathbf{X}_i^q - \mathbf{X}_j^q)^T (\mathbf{X}_i^q - \mathbf{X}_j^q)
 \end{aligned} \tag{6-70}$$

在式 (6-70) 下, 式 (6-68) 和式 (6-69) 变为

$$\begin{aligned}
 \mathbf{S}_{\text{fw}} &= \frac{1}{2} \sum_{q=1}^{\text{row}} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{X}_i^q - \mathbf{X}_j^q)^T (\mathbf{X}_i^q - \mathbf{X}_j^q) \\
 &= \sum_{q=1}^{\text{row}} \mathbf{X}^q (\mathbf{D} - \mathbf{W}) (\mathbf{X}^q)^T
 \end{aligned} \tag{6-71}$$

$$\begin{aligned}
 \mathbf{S}_{\text{fb}} &= \frac{1}{2} \sum_{q=1}^{\text{row}} \sum_{i,j=1}^n \mathbf{W}_{ij}^p (\mathbf{X}_i^q - \mathbf{X}_j^q)^T (\mathbf{X}_i^q - \mathbf{X}_j^q) \\
 &= \sum_{q=1}^{\text{row}} \mathbf{X}^q (\mathbf{D}^p - \mathbf{W}^p) (\mathbf{X}^q)^T
 \end{aligned} \tag{6-72}$$

其中,  $\mathbf{X}^q = [(\mathbf{X}_1^q)^T, (\mathbf{X}_2^q)^T, \dots, (\mathbf{X}_n^q)^T]^T \in \mathbf{R}^{\text{col} \times n}$  由所有样本图像的第  $q$  行组成。

## 2) A2DFLDA

与 2DFLDA 不同, A2DFLDA 用于提取图像的列内信息, 其形式化描述如下: 假设有  $n$  个大小为  $\text{row} \times \text{col}$  的训练样本图像  $\mathbf{X}_i$  ( $i=1, 2, \dots, n$ ), 寻找最优投影矩阵  $\mathbf{V} \in \mathbf{R}^{\text{row} \times d}$  使得  $\mathbf{X}_i$  的低维投影  $\mathbf{Y}_i = \mathbf{V}^T \mathbf{X}_i$ ,  $\mathbf{Y}_i \in \mathbf{R}^{d \times \text{col}}$  ( $d < \text{row}$ ), 使得在低维空间内模糊类内离差度最小的同时模糊类间离差度最大。在图嵌入框架下, A2DFLDA 的模糊类内、类间离差矩阵  $\mathbf{S}_{\text{fw}}$  和  $\mathbf{S}_{\text{fb}}$  分别定义为

$$\mathbf{S}_{\text{fw}} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \tag{6-73}$$

$$\mathbf{S}_{\text{fb}} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{ij}^p (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \tag{6-74}$$

其中,  $\mathbf{W}$  和  $\mathbf{W}^p$  为 FLDA 在图嵌入框架下的权值矩阵。

对于任意  $\mathbf{X}_i, \mathbf{X}_j \in \mathbf{R}^{\text{row} \times \text{col}}$ , 令  $\mathbf{X}_i^q$  表示  $\mathbf{X}_i$  的第  $q$  列, 则有

$$\begin{aligned}
 (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T &= \left( [\mathbf{X}_i^1, \dots, \mathbf{X}_i^{\text{col}}] - [\mathbf{X}_j^1, \dots, \mathbf{X}_j^{\text{col}}] \right) \left( [\mathbf{X}_i^1, \dots, \mathbf{X}_i^{\text{col}}] - [\mathbf{X}_j^1, \dots, \mathbf{X}_j^{\text{col}}] \right)^T \\
 &= \sum_{q=1}^{\text{col}} (\mathbf{X}_i^q - \mathbf{X}_j^q)(\mathbf{X}_i^q - \mathbf{X}_j^q)^T
 \end{aligned} \tag{6-75}$$

在式 (6-75) 下, 式 (6-73) 和式 (6-74) 变为

$$\begin{aligned} \mathbf{S}_{\text{fw}} &= \frac{1}{2} \sum_{q=1}^{\text{col}} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{X}_i^q - \mathbf{X}_j^q)(\mathbf{X}_i^q - \mathbf{X}_j^q)^{\text{T}} \\ &= \sum_{q=1}^{\text{col}} \mathbf{X}^q (\mathbf{D} - \mathbf{W})(\mathbf{X}^q)^{\text{T}} \end{aligned} \quad (6-76)$$

$$\begin{aligned} \mathbf{S}_{\text{fb}} &= \frac{1}{2} \sum_{q=1}^{\text{col}} \sum_{i,j=1}^n \mathbf{W}_{ij}^{\text{p}} (\mathbf{X}_i^q - \mathbf{X}_j^q)(\mathbf{X}_i^q - \mathbf{X}_j^q)^{\text{T}} \\ &= \sum_{q=1}^{\text{col}} \mathbf{X}^q (\mathbf{D}^{\text{p}} - \mathbf{W}^{\text{p}})(\mathbf{X}^q)^{\text{T}} \end{aligned} \quad (6-77)$$

其中,  $\mathbf{X}^q = [\mathbf{X}_1^q, \mathbf{X}_2^q, \dots, \mathbf{X}_n^q] \in \mathbf{R}^{\text{row} \times n}$  由所有样本图像的第  $q$  列组成。

### 3) 总结

从式 (6-71) 和式 (6-72) [式 (6-76) 和式 (6-77)] 可知, 在 2DFLDA (A2DFLDA) 中每一行 (列) 可以看作和原始图像类别相同的样本, 因此训练样本从原来的  $n$  增加至  $\text{row} \times n$  ( $\text{col} \times n$ )。在求解式 (6-54) 对应的广义特征向量问题时,  $\mathbf{S}_{\text{fw}}$  为  $\text{col} \times \text{col}$  ( $\text{row} \times \text{row}$ ) 的矩阵, 在大多数情况下  $\text{row} \times n > \text{col}$  ( $\text{col} \times n > \text{row}$ ), 因此在 2DFLDA (A2DFLDA) 中,  $\mathbf{S}_{\text{fw}}$  很少出现奇异的情况, 最优投影矩阵  $\mathbf{V}$  可以由矩阵  $(\mathbf{S}_{\text{fw}})^{-1} \mathbf{S}_{\text{fb}}$  的最大的  $d$  个最大的特征值对应的特征向量组成。

## 6.4.3 对角图像

从 6.4.2 小节中对 2DFLDA (A2DFLDA) 的分析可知, 图像中的每一行 (列) 可以看作和原来图像类别相同的样本, 每一行 (列) 只反映了行 (列) 内信息的变化而忽略了列 (行) 内信息的变化, 因此 2DFLDA (A2DFLDA) 只是提取了行 (列) 内的特征信息。相较于原始的图像矩阵, 对角图像每一行或列能够很好地同时反映行列信息的变化, 下面介绍对角图像的定义。

对于一幅用  $\text{row} \times \text{col}$  矩阵表示的图像  $\mathbf{I}$ , 其对应的对角图像  $\mathbf{G}$  定义如下。

(1) 如果行数  $\text{row}$  小于列数  $\text{col}$ , 利用图 6.20 (a) 中的方法得到原始图像  $\mathbf{I}$  的对角图像  $\mathbf{G}$ 。

(2) 如果行数  $\text{row}$  大于或等于列数  $\text{col}$ , 利用图 6.20 (b) 中的方法得到原始图像  $\mathbf{I}$  的对角图像  $\mathbf{G}$ 。

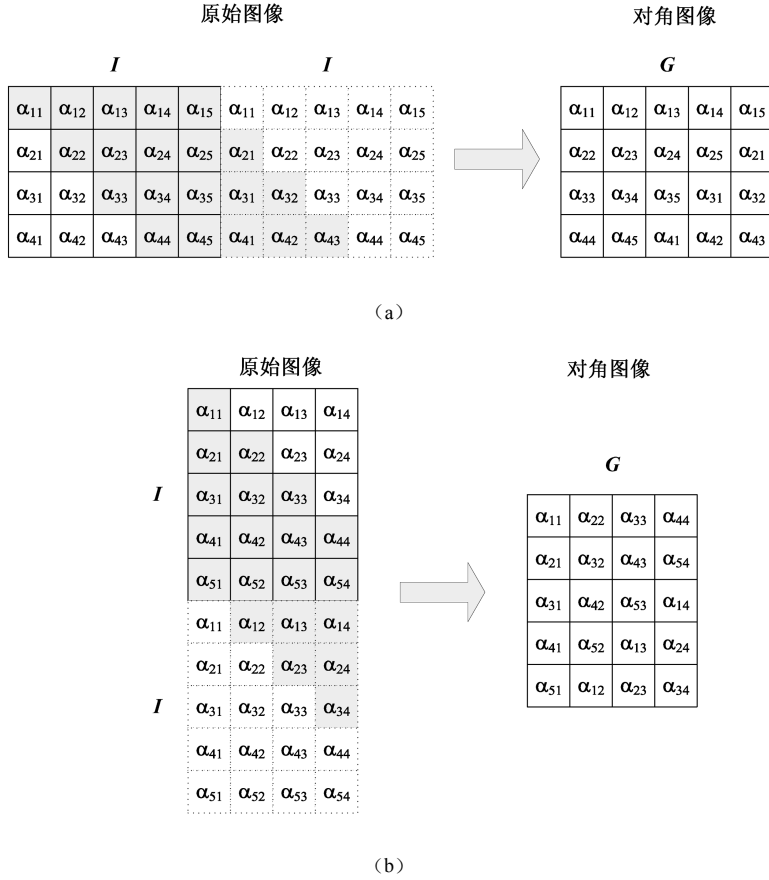


图 6.20 对角图像的产生方法

从图 6.20 中可以看出，利用图 6.20 (a) 得到的对角图像的每一列能够很好地同时反映行列信息的变化，而利用图 6.20 (b) 得到的对角图像的每一行能够很好地同时反映行列信息的变化。在后面的讨论中，我们假设图像的行数  $\text{row}$  大于或等于列数  $\text{col}$ ，即利用图 6.20 (b) 得到对角图像。

#### 6.4.4 DiaFLDA

本小节将对角图像和 FLDA 相结合提出了 DiaFLDA，其形式化描述如下：对于  $n$  个大小为  $\text{row} \times \text{col}$  的训练样本图像  $X_i (i=1, 2, \dots, n)$ ，假设  $\text{row} \geq \text{col}$ ，寻找最优投影矩阵  $V \in \mathbf{R}^{\text{col} \times d}$  使得  $X_i$  的低维投影  $Y_i = X_i V$ ， $Y_i \in \mathbf{R}^{\text{row} \times d} (d < \text{col})$ ，使得在低维空间内模糊对角类内离差度最小的同时模糊对角类间离差度最大。DiaFLDA 算法主要分为 3 步：①计算模糊隶属度矩阵；②将图像转换为对角图像，构造模糊对角类内及类间离差矩

阵；③通过最大化模糊对角类间离差矩阵与模糊对角类内离差矩阵的迹比值得到最优投影矩阵。

### 1. 计算模糊隶属度矩阵

这里采用文献[71]提出的模糊  $k$  最近邻 (FKNN) 方法计算模糊隶属度矩阵。FKNN 方法的计算步骤如下。

(1) 计算训练样本图像集中任意两个样本图像之间的欧式距离，得到一个欧式距离矩阵。

(2) 将欧式距离矩阵的主对角线上的元素的值设置为无穷大。

(3) 将欧式距离矩阵的每一列按照升序顺序进行排列，找到每一个样本的  $k$  个最近邻点。

(4) 利用式 (6-78) 计算样本  $\mathbf{X}_i$  对于类别  $l$  的隶属度。

$$u_{li} = \begin{cases} 0.51 + 0.49 \times (n_{li}/k), & l = l_i \\ 0.49 \times (n_{li}/k), & l \neq l_i \end{cases} \quad (6-78)$$

其中， $n_{li}$  表示在样本  $\mathbf{X}_i$  的  $k$  个最近邻点中类别为  $l$  的个数。模糊隶属度矩阵  $\mathbf{U}$  可以表示为

$$\mathbf{U} = [u_{li}], \quad l = 1, 2, \dots, c, \quad i = 1, 2, \dots, n \quad (6-79)$$

根据模糊隶属度矩阵  $\mathbf{U}$ ，利用推论 6-1 得到图嵌入框架下对应的权值矩阵  $\mathbf{W}$  和  $\mathbf{W}^p$ 。

### 2. 构造模糊对角类内及类间离差矩阵

对于训练集中的每一个样本图像  $\mathbf{X}_i$ ，假设  $\text{row} \geq \text{col}$ ，利用图 6.20 (b) 中的方法得到其对应的对角图像  $\mathbf{G}_i$ ，由 6.4.3 小节的分析可知， $\mathbf{G}_i$  中的每一行能够很好地同时反映行列信息的变化，在图嵌入框架下，利用对角图像矩阵  $\mathbf{G}_i$  替代式 (6-71) 和式 (6-72) 定义的模糊类内及类间离差矩阵中的图像矩阵  $\mathbf{X}_i$  得到模糊对角类内离差矩阵  $\tilde{\mathbf{S}}_{\text{fw}}$  和模糊对角类间离差矩阵  $\tilde{\mathbf{S}}_{\text{fb}}$ ，即

$$\begin{aligned} \tilde{\mathbf{S}}_{\text{fw}} &= \frac{1}{2} \sum_{q=1}^{\text{row}} \sum_{i,j=1}^n \mathbf{W}_{ij} (\mathbf{G}_i^q - \mathbf{G}_j^q)^T (\mathbf{G}_i^q - \mathbf{G}_j^q) \\ &= \sum_{q=1}^{\text{row}} \mathbf{G}^q (\mathbf{D} - \mathbf{W}) (\mathbf{G}^q)^T \end{aligned} \quad (6-80)$$

$$\begin{aligned} \tilde{\mathbf{S}}_{\text{fb}} &= \frac{1}{2} \sum_{q=1}^{\text{row}} \sum_{i,j=1}^n \mathbf{W}_{ij}^p (\mathbf{G}_i^q - \mathbf{G}_j^q)^T (\mathbf{G}_i^q - \mathbf{G}_j^q) \\ &= \sum_{q=1}^{\text{row}} \mathbf{G}^q (\mathbf{D}^p - \mathbf{W}^p) (\mathbf{G}^q)^T \end{aligned} \quad (6-81)$$

其中， $\mathbf{G}^q = [(\mathbf{G}_1^q)^T, (\mathbf{G}_2^q)^T, \dots, (\mathbf{G}_n^q)^T]^T \in \mathbf{R}^{\text{col} \times n}$  由所有对角图像的第  $q$  行组成。

### 3. 最优低维嵌入

最优投影矩阵  $\mathbf{V} \in \mathbf{R}^{\text{col} \times d}$  ( $d < \text{col}$ ) 使得在低维空间内模糊对角类间离差矩阵与模糊对角类内离差矩阵的迹比值最大化, 从式 (6-80) 和式 (6-81) 可知,  $\tilde{\mathbf{S}}_{\text{fw}}$  与  $\tilde{\mathbf{S}}_{\text{fb}}$  均为  $\text{col} \times \text{col}$  的矩阵。与 2DFLDA 一样, 通常情况下  $\tilde{\mathbf{S}}_{\text{fw}}$  为非奇异矩阵, 所以最优投影矩阵  $\mathbf{V}$  由矩阵  $(\tilde{\mathbf{S}}_{\text{fw}})^{-1} \tilde{\mathbf{S}}_{\text{fb}}$  的最大的  $d$  个特征值对应的特征向量组成。

设最优投影矩阵  $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$ , 对于大小为  $\text{row} \times \text{col}$  的测试图像  $\mathbf{X}$  经投影矩阵  $\mathbf{V}$  得到一个大小为  $\text{row} \times d$  的低维投影图像  $\mathbf{Z}$ , 即

$$\mathbf{Z} = \mathbf{XV} \quad (6-82)$$

根据以上分析, DiaFLDA 算法的流程如下。

**输入:**  $c$  个类别  $n$  个样本的训练集  $\{\mathbf{X}_i | i = 1, 2, \dots, n, \mathbf{X}_i \in \mathbf{R}^{\text{row} \times \text{col}}\}$ , 假设  $\text{row} \geq \text{col}$ 。

**执行过程:**

(1) 利用 FKNN 方法计算模糊隶属度矩阵  $\mathbf{U}$ , 并根据推论 6-1 计算图嵌入框架下对应的权值矩阵  $\mathbf{W}$  和  $\mathbf{W}^p$ 。

(2) 根据 6.4.3 小节介绍的方法得到训练集每一幅图像  $\mathbf{X}_i$  的对角图像  $\mathbf{G}_i$ , 并根据式 (6-80) 和式 (6-81) 计算模糊对角类内离差矩阵  $\tilde{\mathbf{S}}_{\text{fw}}$  和模糊对角类间离差矩阵  $\tilde{\mathbf{S}}_{\text{fb}}$ 。

(3) 计算矩阵  $(\tilde{\mathbf{S}}_{\text{fw}})^{-1} \tilde{\mathbf{S}}_{\text{fb}}$  的最大的  $d$  个特征值对应的特征向量, 并由其组成最优投影矩阵  $\mathbf{V}$ 。

**输出:** 最优投影矩阵  $\mathbf{V}$ 。

### 4. 两阶段 DiaFLDA

在 DiaFLDA 中, 图像矩阵  $\mathbf{X}_i$  经投影矩阵  $\mathbf{V} \in \mathbf{R}^{\text{col} \times d}$  得到的低维图像矩阵  $\mathbf{Y}_i$  的大小为  $\text{row} \times d$ , 低维图像的列数减少而行数没有发生变化, 这样用于表示图像特征的矩阵仍然占用巨大的存储空间, 为此我们提出了两阶段 DiaFLDA (Two Stage DiaFLDA, TDiaFLDA), 利用 DiaFLDA 和 A2DFLDA 进行特征提取。

假设  $\mathbf{V}_1 \in \mathbf{R}^{\text{col} \times d_1}$  ( $d_1 < \text{col}$ ) 为用 DiaFLDA 得到的最优投影矩阵,  $\mathbf{V}_2 \in \mathbf{R}^{\text{row} \times d_2}$  ( $d_2 < \text{row}$ ) 为用 A2DFLDA 得到的最优投影矩阵, 则图像矩阵  $\mathbf{X}_i$  经  $\mathbf{V}_1$  和  $\mathbf{V}_2$  得到的低维图像矩阵  $\mathbf{Y}_i = \mathbf{V}_2^T \mathbf{X}_i \mathbf{V}_1$  的大小为  $d_2 \times d_1$ , 可以大大减少表示图像特征的矩阵所占有的存储空间。

对于一个新的图像矩阵  $\mathbf{X}$ , 其低维表示为  $\mathbf{Z} = \mathbf{V}_2^T \mathbf{XV}_1$ , 利用最近邻分类器确定图像  $\mathbf{X}$  所属类别。



## 6.5 DCT 域内拉普拉斯值排序的子空间特征提取方法

### 6.5.1 方法提出的背景

传统的线性子空间特征提取方法是一种基于向量的特征提取方法,在应用该类方法时,经常遭遇小样本问题。为此,通常先利用 PCA 方法去除较小的主元成分从而将图像向量转换到 PCA 子空间,然后在 PCA 子空间内执行相应的特征提取方法,这样就会丢失一些有用的鉴别信息,同时,执行 PCA 计算复杂度较高。离散余弦变换(DCT)是一种有效的图像压缩和识别算法<sup>[72,73]</sup>。与 PCA 相比, DCT 的优点在于数据是独立的,也就是说,基图像仅仅依赖于图像本身而不是整个图像集;另外,我们可以利用快速傅里叶变换实现 DCT,从而大大提高计算的效率。有研究表明 PCA、LDA 算法在 DCT 域内可以得到更好的识别效果<sup>[74,75]</sup>,受其启发,文献[76]提出了一种 DCT 与 LPP 相结合的特征提取算法(DCT+LPP 算法)。在该算法中,利用 DCT 代替 PCA 进行降维,然后在低维空间中利用 LPP 进行特征提取。在基于 DCT 的人脸识别方法中, DCT 系数的数量与所取得的识别率并不成正比,因此如何选择最有效的 DCT 系数作为识别特征是这类算法的关键问题。现有的基于 DCT 的人脸识别方法都是按矩形或 Z 字形顺序选择低频 DCT 系数作为特征进行识别的<sup>[77]</sup>。拉普拉斯值(Laplacian Score, LS)是一种有效的特征选择算法,主要用于评价特征的局部保持能力。本节从有效特征选择角度出发,以基于 LS 的特征选择算法作为 DCT 系数选择的依据,提出了 DCT 域内拉普拉斯值排序的子空间特征提取方法。

### 6.5.2 离散余弦变换(DCT)

DCT 是由 N.Ahmed 等人提出的一种与傅里叶变换(FFT)紧密相关的正交变换,在语音、图像等数据压缩领域得到了很好的应用。由于 DCT 是正交变换,因此在 DCT 域内和在原始空间上执行 LPP 算法其结果相同。与 PCA 相比, DCT 是基于单个样本的,而不是整个样本集,所以当有新的训练样本加入时,仅需要对新加入的样本进行 DCT,同时可以利用快速 FFT 实现 DCT,可以大大降低算法的计算复杂度。因此,在 DCT+LPP 算法中可以利用 DCT 代替 PCA 进行降维,然后在 DCT 域内执行 LPP。

#### 1. DCT 概述

DCT 的详细描述可参阅第 3 章。这里主要讨论二维的情况,对于一个  $M \times N$  的二维矩阵  $X$  的 DCT 结果可以通过在行方向和列方向上进行 DCT 得到,即

$$Y = C_M^T X C_N, \quad X = C_M Y C_N^T \quad (6-83)$$

其中,  $C_M \in \mathbf{R}^{M \times M}$  和  $C_N \in \mathbf{R}^{N \times N}$  分别为行方向和列方向上的正交变换矩阵。

二维矩阵  $X$  的 DCT 结果为一个与  $X$  大小相同的矩阵, 称为矩阵  $X$  的 DCT 系数矩阵。图 6.21 为一幅人脸图像及其 DCT 系数矩阵。

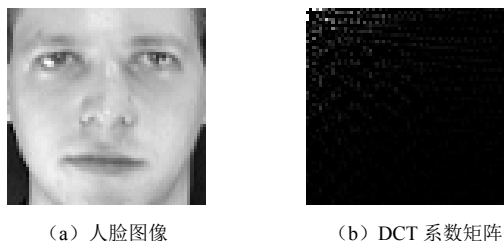


图 6.21 人脸图像及其 DCT 系数矩阵

图 6.21 (b) 的 DCT 系数矩阵很好地体现了其“能量集中”这一特性, 即 DCT 系数矩阵的左上角的数值较大, 说明人脸图像主体信息主要集中在 DCT 系数的低频部分。

在应用 LPP 算法之前, 需要将  $M \times N$  的图像矩阵转换为  $MN$  维向量, 式 (6-83) 可以转换为如式 (6-84) 所示的向量形式。

$$\hat{Y} = G^T \hat{X}, \quad \hat{X} = G \hat{Y} \quad (6-84)$$

其中,  $\hat{X} = [x_{0,0}, \dots, x_{M-1,N-1}]$  和  $\hat{Y} = [y_{0,0}, \dots, y_{M-1,N-1}]$  为  $MN$  维向量;  $G$  为  $MN \times MN$  的正交变换矩阵, 其值为

$$G = \begin{bmatrix} c_{0,0}c_{0,0} & \cdots & c_{0,M-1}c_{0,0} & \cdots & \cdots & c_{0,M-1}c_{0,N-1} \\ \vdots & & \vdots & & & \vdots \\ c_{M-1,0}c_{0,0} & \cdots & c_{M-1,M-1}c_{0,0} & \cdots & \cdots & c_{M-1,M-1}c_{0,N-1} \\ \vdots & & \vdots & & & \vdots \\ \vdots & & \vdots & & & \vdots \\ c_{M-1,0}c_{N-1,0} & \cdots & c_{M-1,M-1}c_{N-1,0} & & & c_{M-1,M-1}c_{N-1,N-1} \end{bmatrix} \quad (6-85)$$

此外, 向量  $\hat{Y}$  中元素的序列对应于变换矩阵  $G$  的列顺序, 因此  $\hat{Y}$  中元素序列的改变不会改变变换矩阵  $G$  的正交性。

## 2. 分块 DCT

对于 JPEG 压缩标准, 首先将图像划分为若干个大小为  $8 \times 8$  的子块 (子图像), 然后再对各个子块分别执行 DCT。同样, 对于一个大小为  $pn \times qn$  的图像, 若划分为  $pq$  个大小为  $n \times n$  的子块, 则其分块 DCT 可以表示为

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{p1} \\ \vdots \\ Y_{pq} \end{bmatrix} = \begin{bmatrix} G_{11} & & & \\ & \ddots & & \\ & & G_{p1} & \\ & & & \ddots \\ & & & & G_{pq} \end{bmatrix} \times \begin{bmatrix} X_{11} \\ \vdots \\ X_{p1} \\ \vdots \\ X_{pq} \end{bmatrix} \quad (6-86)$$

其中,  $G_{ij}$  为对每一个子块进行 DCT 时对应的变换矩阵。由于  $G_{ij}$  为正交矩阵, 即有  $G_{ij}^T = G_{ij}^{-1}$ , 所以有

$$\begin{bmatrix} G_{11} & & & \\ & \ddots & & \\ & & G_{p1} & \\ & & & \ddots \\ & & & & G_{pq} \end{bmatrix}^T = \begin{bmatrix} G_{11} & & & \\ & \ddots & & \\ & & G_{p1} & \\ & & & \ddots \\ & & & & G_{pq} \end{bmatrix}^{-1} \quad (6-87)$$

因此, 式 (6-87) 对应的块对角变换矩阵同样也为正交矩阵, 所以分块 DCT 系数可以直接应用 LPP 算法。

### 6.5.3 局部保持能力判据

拉普拉斯值 (LS) 作为局部保持能力判据选择能更好地刻画样本流形结构的 DCT 系数, 其本质上与 LPP 相似<sup>[78,79]</sup>。令  $f_{ri}$  为第  $i$  个样本  $\mathbf{x}_i$  的第  $r$  个特征,  $i=1,2,\dots,N$ , LS 的计算过程如下。

(1) 构造近邻图  $G$ : 如果样本  $\mathbf{x}_i$  与样本  $\mathbf{x}_j$  互为近邻点, 那么  $\mathbf{x}_i$  与  $\mathbf{x}_j$  之间有一条边相连。

(2) 权值矩阵: 如果  $\mathbf{x}_i$  与  $\mathbf{x}_j$  有一条边相连, 则  $W_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ , 其中  $\sigma$  为经验值, 否则  $W_{ij} = 0$ 。

(3) 对于第  $r$  个特征  $\mathbf{f}_r = [f_{r1}, f_{r2}, \dots, f_{rN}]^T$ , 其 LS 定义为

$$L_r = \frac{\sum_{ij} (f_{ri} - f_{rj}) W_{ij}}{\text{Var}(\mathbf{f}_r)} \quad (6-88)$$

其中,  $\text{Var}(\mathbf{f}_r)$  为第  $r$  个特征的方差, 经过简单变形, 式 (6-88) 变为

$$L_r = \frac{\tilde{\mathbf{f}}_r^T \mathbf{L} \tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^T \mathbf{D} \tilde{\mathbf{f}}_r} \quad (6-89)$$

其中,  $\tilde{f}_r = f_r - \frac{f_r^T D \mathbf{1}}{\mathbf{1}^T D \mathbf{1}} \mathbf{1}$ ;  $D = \sum_j W_{ij}$ ;  $L = D - W$ ;  $\mathbf{1} = [1, \dots, 1]^T \in \mathbf{R}^N$ 。

根据式 (6-88) 中 LS 的定义, 一个好的特征, 应该使得  $\sum_{ij} (f_{ri} - f_{rj}) W_{ij}$  最小化,  $\text{Var}(f_r)$  最大化, LS 趋向取较小的值。  $\sum_{ij} (f_{ri} - f_{rj}) W_{ij}$  最小化表明该特征具有较强的局部信息保持能力, 即互为近邻点的两个样本点在该特征上差别最小;  $\text{Var}(f_r)$  最大化表明该特征具有较强的样本表示能力。因此, 可以将 LS 作为局部保持能力判据, 其值越小, 表明该特征刻画样本流形结构的能力越强。同样, 在执行 LPP 算法之前, 将分块 DCT 得到的分块 DCT 系数按照 LS 从小到大的顺序转换为一维向量形式, 如图 6.22 所示。

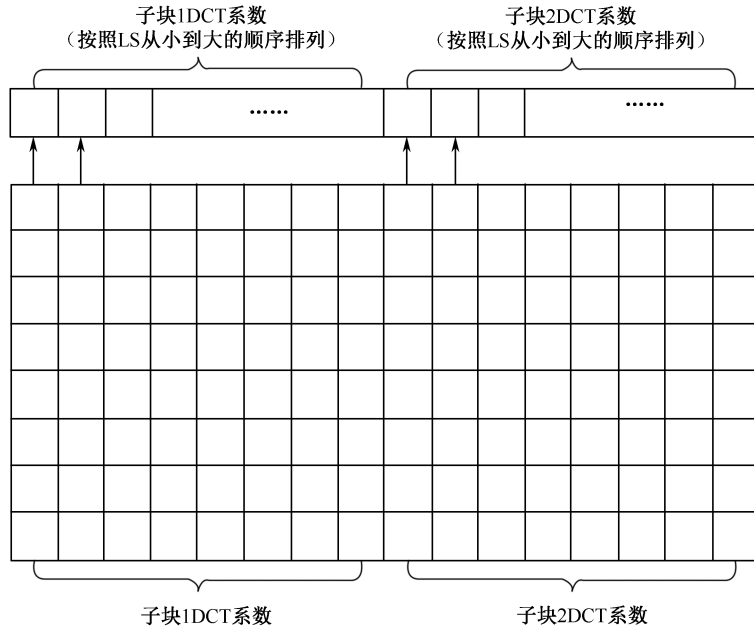


图 6.22 基于 LS 的 DCT 系数选择

从 ORL 人脸库中  $(32 \times 32)^{[80]}$  随机选取每一个人的 5 幅图像组成训练样本集, 剩余的图像组成测试样本集。首先将训练样本集中每一幅图像划分为  $8 \times 8$  的子图像, 这样每一幅图像可以得到 16 幅子图像, 然后对每一幅子图像进行 DCT, 每一幅子图像对应 64 个 DCT 系数。图 6.23 为每一幅子图像从上到下从左到右对应的 DCT 系数的 LS。

从图 6.23 中可以看出, 从低频 DCT 系数到高频 DCT 系数, 其对应的 LS 的变化并不是单调递增的, 而呈现出一种振荡式变化趋势, 也就是说, 低频的 DCT 系数的局部保持能力不一定低于高频的 DCT 系数。因此, 采用矩形或 Z 字形对 DCT 系数进行选择, 并不能将具有较强局部保持能力的 DCT 系数选择出来。这里根据 DCT 系数的 LS 按照从小到大的顺序进行选择。

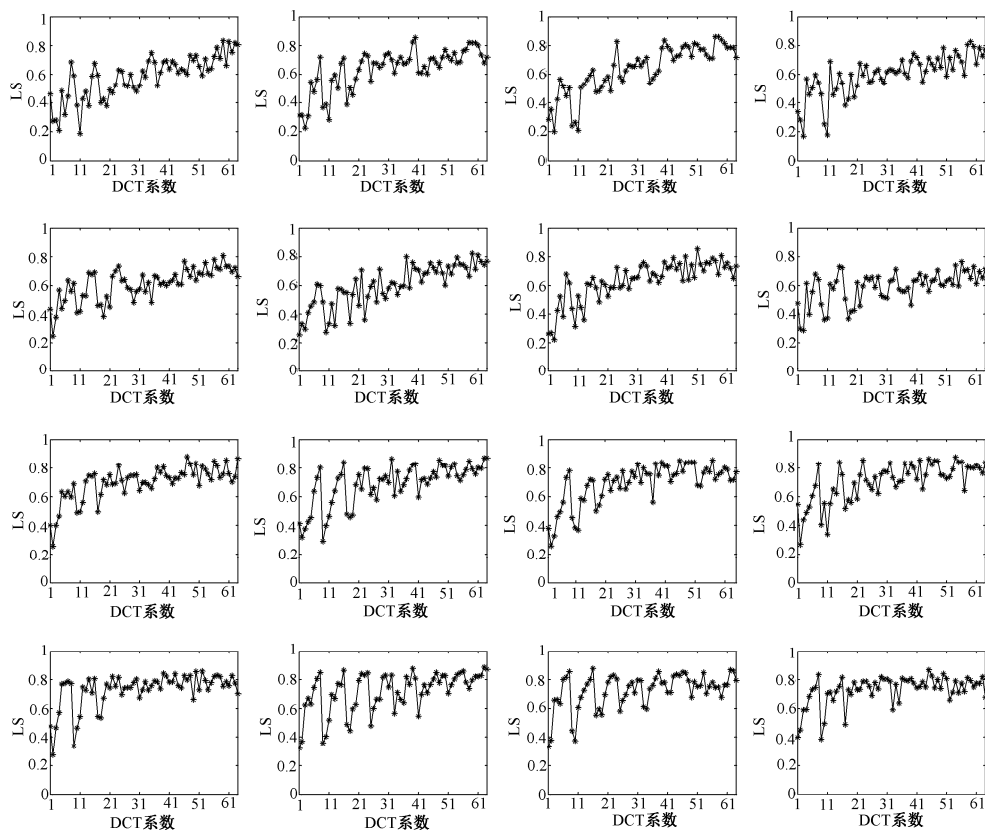


图 6.23 子图像的 DCT 系数对应的 LS

#### 6.5.4 DCT/LS+LPP

DCT/LS+LPP 算法的流程图如图 6.24 所示, 包括两个阶段: 训练阶段和识别阶段。

在训练阶段, 首先对训练集中每一幅图像划分成若干个大小为  $n \times n$  的子图像; 然后对各个子块进行 DCT, 得到分块 DCT 系数; 其次对于每一个子块 DCT 系数, 在不同频率的 DCT 系数上计算其 LS 作为局部保持能力判据, 按 LS 从小到大进行排序; 最后将每一个子块 DCT 系数中 LS 较小的 DCT 系数组合成一个一维向量作为图像的特征执行 LPP 算法, 得到最优投影矩阵和训练样本的识别特征。

在识别阶段, 对于一幅测试图像, 同样首先将其划分成若干个大小为  $n \times n$  的子图像; 然后对各个子块进行 DCT, 得到分块 DCT 系数; 其次在每一个子块内依据训练阶段中的次序选择 DCT 系数, 并将每一个子块所选取的 DCT 系数组成的一维向量在

训练阶段获得的最优投影矩阵的投影结果作为图像的识别特征；最后采用利用欧式距离作为相似度度量的最近邻分类器完成对测试图像的分类。

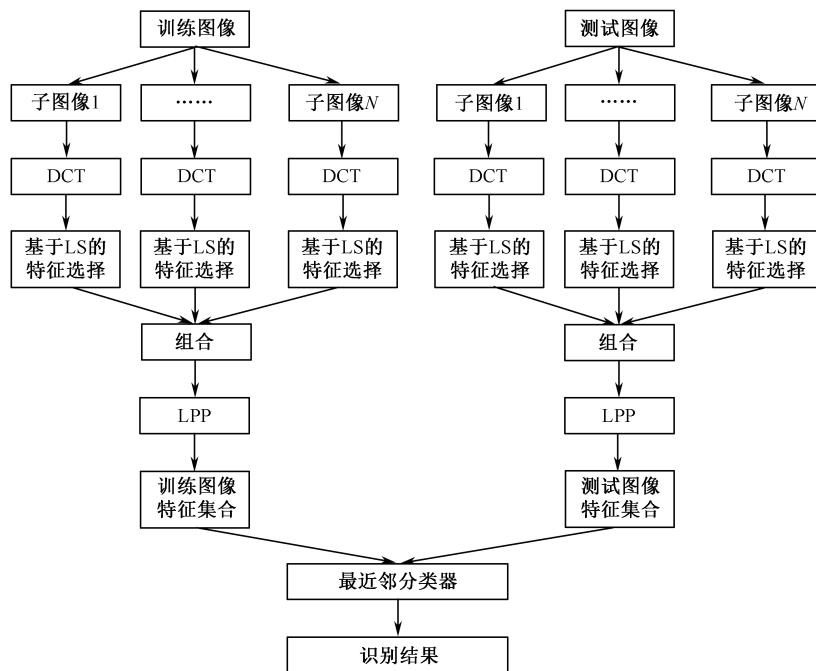


图 6.24 DCT/LS+LPP 算法的流程图

## 参 考 文 献

- [1] Grigorios T. Dimensionality reduction and sparse representations in computer vision[D]. Rochester: Rochester Institute of Technology, 2011.
- [2] Bishop C M. Pattern recognition and machine learning[M]. New York: Springer, 2006.
- [3] He X F. Locality preserving projection[D]. Chicago: Chicago University, 2005.
- [4] 王永茂. 子空间特征提取方法及其应用研究[D]. 北京: 北京科技大学, 2013.
- [5] Jolliffe I T. Principal component analysis[D]. Berlin: Springer, 2002.
- [6] Martinez A, Kak A. PCA versus LDA[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2):228-233.
- [7] 杨琼, 丁晓青. 对称主分量分析及其在人脸识别中的应用[J]. 计算机学报, 2003, 26(9):1146-1151.
- [8] Gottumukkal R, Asari V K. An improved face recognition technique based on modular

- PCA approach[J]. Pattern Recognition Letter, 2004, 25(4):429-436.
- [9] Chen S C, Zhu Y L. Subpattern-based principle component analysis[J]. Pattern Recognition, 2004, 37(1):1081-1083.
- [10] Tan K, Chen S C. Adaptively weighted sub-pattern PCA for face recognition[J]. Neurocomputing, 2005, 64(1-4):205-511.
- [11] Swets D L, Weng J Y. Using discriminant eigenfeatures for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8):831-836.
- [12] Yu H, Yang J. A direct LDA algorithm for high-dimensional data with application to face recognition[J]. Pattern Recognition, 2001, 34(10): 2067-2070.
- [13] Song F X, Zhang D, Wang J Z, et al. A parameterized direct LDA and its application to face recognition[J]. Neurocomputing, 2007, 71(1-3):191-196.
- [14] Zhou D, Yang X. Face recognition using direct-weighted LDA[C] // Proceedings of 8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand, 2004:760-768.
- [15] Chen L F, Liao H Y, Ko M T, et al. A new LDA-based face recognition system which can solve the small samples size problem[J]. Pattern Recognition, 2000, 33(10): 1713-1726.
- [16] Wang X G, Tang X O. Dual-space linear discriminant analysis for face recognition[C] // Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Washington, DC, United States, 2004:564-569.
- [17] Friedman J H. Regularized discriminant analysis[J]. Journal of the American Statistical Association, 1989, 84(405):165-175.
- [18] Howland P, Park H. Generalizing discriminant analysis using the generalized singular value decomposition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(8):995-1006.
- [19] Ye J P, Janardan R, Park C H, et al. An optimization criterion for generalized discriminant analysis on undersampled problems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(8):982-994.
- [20] Hotelling H. Relations between two sets of variates[J]. Biometrika, 1936, 28(3):321-377.
- [21] Comon P. Independent component analysis-A new concept[J]. Signal Processing, 1994, 36(3):287-314.
- [22] 洪泉, 陈松灿, 倪雪蕾. 子模式典型相关分析及其在人脸识别中的应用[J]. 自动化学报, 2008, 34(1):21-30.
- [23] Lee D D, Seung H S. Algorithms for non-negative matrix factorization[C] // Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada,

- 2001:556-562.
- [24] Vapnik V N. Statistical learning theory[M]. New York: Wiley, 1998.
  - [25] Muler K R, Mika S, Rasch G, et al. An introduction to kernel-based learning algorithms[J]. IEEE Transactions on Neural Networks, 2001, 12(3):181-201.
  - [26] 陈文安. 子空间方法及其核扩展的研究[D]. 北京: 北方工业大学, 2006.
  - [27] Scholkopf B, Smola G, Muller K R. Nonlinear component analysis as kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5):1299-1319.
  - [28] Mika S, Ratsch G, Weston J. Fisher discriminant analysis with kernels[C] // Proceedings of IEEE International Workshop on Neural Networks for Signal Processing, Madison, USA, 1999:41-48.
  - [29] Bach F R, Jordan M I. Kernel independent component analysis[J]. Journal of Machine Learning Research, 2002, 3(1):1-48.
  - [30] Melzer T. Generalized canonical correlation analysis for object recognition[D]. Wien: Vienna University of Technology, 2002.
  - [31] 赵松, 潘可, 张培仁. DLLE: 一种姿态无关的人脸识别改进算法[J]. 小型微型计算机系统, 2009, 30(6):1193-1197.
  - [32] 罗四维, 赵连伟. 基于谱图理论的流形学习算法[J]. 计算机研究与发展, 2006, 43(7):1173-1179.
  - [33] Tenenbaum J, Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500):2319-2323.
  - [34] Griffiths T L, Kalish M L. A multidimensional scaling approach to mental multiplication[J]. Memory and Cognition, 2002, 30(1):97-106.
  - [35] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500):2323-2326.
  - [36] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6):1373-1396.
  - [37] He X F, Niyogi P. Locality preserving projections[C] // Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2003:153-160.
  - [38] Song Y Q, Nie F P, Zhang C S. Semi-supervised sub-manifold discriminant analysis[J]. Pattern Recognition Letters, 2008, 29(13):1806-1813.
  - [39] Cai D, He X F, Han J W. Semi-supervised discriminant analysis[C] // Proceedings of 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007:1-7.
  - [40] Sugiyama M, Id T, Nakajima S, et al. Semi-supervised local fisher discriminant analysis for dimensionality reduction[J]. Machine Learning, 2010, 78(1-2):35-61.



- [41] Huang H, Li J W, Liu J M. Enhanced semi-supervised local fisher discriminant analysis for face recognition[J]. Future Generation Computer Systems, 2012, 28(1):244-253.
- [42] 陈诗国, 张道强. 半监督降维方法的实验比较[J]. 软件学报, 2011, 2(1):28-43.
- [43] Ratthachat R, Kijirikul B. A unified semi-supervised dimensionality reduction framework for manifold learning[J]. Neurocomputing, 2010, 73(10-12):1631-1640.
- [44] Song Y Q, Nie F P, Zhang C S. A unified framework for semi-supervised dimensionality reduction[J]. Pattern Recognition, 2008, 41(9):2789-2799.
- [45] Yang J, Zhang D, Frangi A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(1):131-137.
- [46] Xiong H L, Swamy M N S, Ahmad M O. Two-dimensional FLD for face recognition[J]. Pattern Recognition, 2005, 38(7):1121-1124.
- [47] Zhang D Q, Zhou Z H.  $(2D)^2$  PCA: 2-directional 2-dimensional PCA for efficient face representation and recognition[J]. Neurocomputing, 2005, 69(1-3):224-231.
- [48] Nagabhushan P, Guru D S, Shekar B H.  $(2D)^2$  FLD: An efficient approach for appearance based object recognition[J]. Neurocomputing, 2006, 69(7-9):934-940.
- [49] He X F, Deng C, Partha N. Tensor subspace analysis[C] //Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2005:499-506.
- [50] Lu H P, Plataniotis K N, Venetsanopoulos A N. A survey of multilinear subspace learning for tensor data[J]. Pattern Recognition, 2011, 44(7): 1540-1551.
- [51] 徐东. 数据降维算法的研究及其应用[D]. 合肥: 中国科学技术大学, 2005.
- [52] He X F, Yan S C, Hu Y, et al. Learning a locality preserving subspace for visual recognition[C] // Proceedings of IEEE International Conference on Computer Vision, Beijing, 2003:385-392.
- [53] Lathauwer L D. Signal processing based on multilinear algebra[D]. Leuven: Katholieke Universiteit Leuven, 1997.
- [54] Nie F P, Xiang S M, Song Y Q, et al. Extracting the optimal dimensionality for local tensor discriminant analysis[J]. Pattern Recognition, 2009, 42(1):105-144.
- [55] Yan S C, Xu D, Yan Q, et al. Multilinear discriminant analysis for face recognition[J]. IEEE Transactions on Image Processing, 2007, 16(1):212-220.
- [56] Yan S C, Xu D, Zhang B Y, et al. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1):40-51.

- [57] Zhao J D, Lu K, He X F. Locality sensitive semi-supervised feature selection[J]. Neurocomputing, 2008, 71(10-12):1842-1849.
- [58] Chen H T, Chang H W, Liu T L. Local discriminant embedding and its variants[C] // Proceedings of International Conference on Computer Vision and Pattern Recognition, San Diego, CA, United States, 2005:846-853.
- [59] Sugiyama M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis[J]. Journal of Machine Learning Research, 2007, 8(5):1027-1061.
- [60] 谢钧, 刘剑. 一种新的局部判别投影方法[J]. 计算机学报, 2011, 34(11):2243-2250.
- [61] Raducanu B, Dornaika F. A supervised non-linear dimensionality reduction approach for manifold learning[J]. Pattern Recognition, 2012, 45(6):2432-2444.
- [62] 王永茂, 徐正光, 赵珊. 基于自适应近邻图嵌入的局部鉴别投影算法[J]. 电子与信息学报, 2013, 35(3):633-638.
- [63] Kwak K C, Pedrycz W. Face recognition using a fuzzy fisherface classifier[J]. Pattern Recognition, 2005, 38(10):1717-1732.
- [64] Zadeh L A. Fuzzy sets[J]. Information and control, 1965, 8:338-353.
- [65] Yang W K, Yan X Y, Zhang L, et al. Feature extraction based on fuzzy 2DLDA[J]. Neurocomputing, 2010, 73(10):1556-1561.
- [66] 林宇生, 房福龙, 杨万扣. 模糊二维线性鉴别分析算法[J]. 无线电工程, 2011, 41(9):15-17.
- [67] Zhang D Q, Zhou Z H, Chen S C. Diagonal principal component analysis for face recognition[J]. Pattern Recognition, 2006, 39(1):140-142.
- [68] Nousath S, Hemantha K G, Shivakumara P. Diagonal Fisher linear discriminant analysis for efficient face recognition[J]. Neurocomputing, 2006, 69(13-15):1711-1716.
- [69] Veerabhadrapa, Rangarajan L. Diagonal and secondary diagonal locality preserving projection for object recognition[J]. Neurocomputing, 2010, 73(16-18):3328-3333.
- [70] Wang Y M, Wang Y K. Diagonal Fuzzy Linear Discriminant Analysis for Face Recognition[J]. Journal of computational information systems, 2013, 9(22):9121-9129.
- [71] Keller J M, Gray M R, Givern J A. A fuzzy k-nearest neighbor algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1985, 15(4):580-585.
- [72] Hafed Z M, Levine M D. Face recognition using the discrete cosine transform[J]. International Journal of Computer Vision, 2001, 43(3):167-188.
- [73] Pan Z J, Rust A G, Bolouri H. Image redundancy reduction for neural network classification using Discrete Cosine Transforms[C] // Proceedings of International Joint

- Conference on Neural Networks, Como, Italy, 2000:149-154.
- [74] Ramasubramanian D, Venkatesh Y V. Encoding and recognition of faces based on the human visual model and DCT[J]. Pattern Recognition, 2001, 34(12):2447-2458.
- [75] 张燕昆, 刘重庆. 一种新颖的基于 LDA 的人脸识别方法[J]. 红外与毫米波学报, 2003, 22(5):327-330.
- [76] Zheng Z L, Zhao J M. Locality preserving projection in orthogonal domain[C] // Proceedings of Congress on Images and Signal, Sanya, 2008:613-617.
- [77] 尹洪涛, 付平, 沙学军. 基于 DCT 和线性判别分析的人脸识别[J]. 电子学报, 2009, 37(10):2211-2214.
- [78] He X F, Deng C, Niyogi P. Laplacian score for feature selection[C] // Advances in Neural Information Processing System, Vancouver, British Columbia, Canada, 2005: 507-514.
- [79] Huang H, Feng H L, Peng C Y. Complete local fisher discriminant analysis with laplacian score ranking for face recognition[J]. Neurocomputing, 2012, 89(7):64-77.
- [80] Olivetti & Oracle Research Laboratory. The Olivetti & Oracle Research Laboratory Face Database[DB]. <http://www.cam-orl.co.uk/faceddatabase.html>, 1994.

## 反侵权盗版声明

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复制、销售或通过信息网络传播本作品的行为；歪曲、篡改、剽窃本作品的行为，均违反《中华人民共和国著作权法》，其行为人应承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人员，并保证举报人的信息不被泄露。

举报电话：(010) 88254396；(010) 88258888

传 真：(010) 88254397

E-mail: dbqq@phei.com.cn

通信地址：北京市万寿路 173 信箱

电子工业出版社总编办公室

邮 编：100036